Università degli Studi di Padova

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Statistiche
Corso di Dottorato di Ricerca in Scienze Statistiche
Ciclo XXXI

Propensity score techniques in multiple treatments framework: the estimation of neighbourhood effect

Coordinatore del Corso: Prof. Nicola Sartori

Supervisore: Prof. Giovanna Boccuzzo

Co-supervisore: Prof. Bruno Arpino

Co-supervisore: Prof. Giuseppe Costa

Dottorando/a: Silan Margherita

Abstract

Neighbourhood effects have been defined by Oakes (2004) as the independent causal effects of neighbourhood on a given number of health or social outcome(s).

The aim of this thesis is to estimate the neighbourhood effect on old population in Turin with a propensity score approach. To achieve this goal, we need to work on adapting propensity score techniques to work well in a framework with many treatments (with ten or more treatments). Data used in the thesis come from the Turin Longitudinal Study (SLT), described in chapter 3. Our main goal is to understand if the observed differences in health outcomes across neighbourhoods can be causally attributed to neighbourhoods' as opposed to their different composition, i.e. to the fact that individuals with different risks factors live in different areas.

In order to adjust for confounders and simulate an experimental approach, we focused on propensity score techniques that are briefly described in chapter 2. The first part of the analysis focuses on the performance evaluation of an inverse probability of treatment approach (IPTW) in a 10-treatment framework (chapter 4) and its application on real data on two different health outcomes: hospitalized fractures and mental health (chapter 5).

In the second part of the thesis we propose a novel method that consists on a matching based on partially ordered sets (poset) (chapter 6). The Matching on Poset based Average Rank for Multiple Treatment (MARMoT), tested with some simulations, has revealed to be really useful to estimate neighbourhood effect, reducing bias of estimates because of the initial improvement of covariates' balance between groups.

Sommario

L'effetto di vicinato è stato definito da Oakes (2004) come l'effetto causale indipendente di un vicinato su qualsiasi esito sociale o di salute. Lo scopo principale di questo elaborato consiste nello stimare l'effetto di vicinato sullo stato di salute degli anziani residenti a Torino con un approccio basato sull'uso del *propensity score*. Tuttavia, risulta necessario adattare le tecniche di *propensity score*, generalmente utilizzate con trattamenti dicotomici, a casi di trattamento multiplo, in cui siano eventualmente coinvolti molti trattamenti (10 o più). I dati utilizzati nella tesi provengono dallo studio longitudinale torinese (SLT), descritto nel capitolo 3. Una delle domande di ricerca principali in questa tesi consiste nello stimare quanto le differenze osservate nello stato di salute degli anziani residenti in diversi vicinati siano dovute al vicinato di residenza e quanto invece siano legate alle diverse caratteristiche degli individui che lo compongono.

Per aggiustare per l'effetto dei confondenti e ricostruire un approccio sperimentale, abbiamo preferito adottare tecniche basate sull'uso del *propensity score*, che sono brevemente descritte nel capitolo 2. Nella prima parte delle analisi viene valutato il funzionamento di un approccio di *inverse probability of treatment weighting* in uno scenario costituito da 10 trattamenti (capitolo 4). Viene poi applicato su dati reali per stimare l'effetto di vicinato su due esiti di salute: le fratture ospedalizzate e le malattie mentali (capitolo 5).

Nella seconda parte della tesi invece viene descritta una proposta originale che consiste in un *matching* che sfrutta la teoria degli ordinamenti parziali (*poset*). Questo approccio, che abbiamo chiamato *Matching on Poset based Average Rank for Multiple Treatment (MARMoT)*, è stato testato attraverso uno studio di simulazione e si è rivelato essere particolarmente utile per la stima degli effetti di vicinato, riducendo la distorsione delle stime grazie al miglioramento del bilanciamento delle variabili confondenti tra i vicinati.

Acknowledgements

Desidero ringraziare tutti coloro che mi hanno sostenuta e accompagnata negli ultimi tre anni.

In primo luogo ringrazio la mia relatrice, la Professoressa Giovanna Boccuzzo, che è stata per me una guida e un punto di riferimento in questo percorso. Sono grata al Servizio Sovrazonale di Epidemiologia di Grugliasco che mi ha accolta nelle prime fasi del mio lavoro, in particolare desidero ringraziare il Professor Giuseppe Costa, mio co-relatore, e tutti i suoi collaboratori. Inoltre, ringrazio di cuore anche il professor Bruno Arpino, mio co-relatore, per i suoi consigli che hanno arricchito immensamente questa tesi e per avermi ospitata presso l'Università Pompeu Fabra a Barcellona.

Desidero ringraziare anche il dipartimento di Scienze Statistiche dell'Università di Padova, che è stato negli ultimi otto anni quasi una seconda casa per me, il direttore del dottorato, Professor Nicola Sartori, il Professor Adriano Paggiaro per la sua disponibilità e i suoi saggi consigli, Patrizia Piacentini e tutti i miei colleghi. Grazie anche a Giulio Caperna che mi ha sempre offerto aiuto e consiglio e che ha condiviso con me il codice R per l'approssimazione dell'average rank.

Non meno importanti in questi ultimi tre anni sono stati famigliari e amici che mi hanno sempre sostenuta, in particolare mio marito Federico che mi è stato accanto con pazienza e amore in ogni momento.

Contents

List of Figures

List of Tables

Introduction

Overview

During last 20 years the interest in effects of neighbourhood context on individuals' lives has grown exponentially (Arcaya *et al.*, 2016), creating a new important research field in social epidemiology. These effects are usually called neighbourhood effects, indeed, they have been defined by Oakes (2004) as the independent causal effects of neighbourhood on a given number of health or social outcome(s).

In this thesis we are going to estimate the neighbourhood effect on two health outcomes on old population of residents in Turin: hospitalized bones fractures and mental health. Indeed, the interest in this topic comes from a real need of the the Unit "SCaDU Servizio Sovrazonale di Epidemiologia" in Grugliasco (Turin, Italy). In some researches old people are shown to be more susceptible to neighbourhood effect, because they spend more time close to the neighbourhood they live with respect to younger and working people (Melis *et al.*, 2015; Turrell *et al.*, 2014).

A key issue when estimating the neighbourhood effect is that its estimation typically relies on observational data. The most valid way to obtain unbiased estimates would be to conduct community trials - because they are designed to control for the treatment allocation mechanisms, and make them largely negligible (Oakes, 2004) - but in practise they are infeasible in neighbourhood effect estimation (Harding, 2003). So the crucial question in most studies on the neighbourhood effect is whether differences between neighbourhoods can be causally attributed to the neighbourhood itself, rather than to differences between individuals living in the various areas (Harding, 2003).

Neighbourhood effects have been typically estimated using multilevel models, where neighbourhoods are represented as the highest level. Regression models can help us to adjust for observed confounders, when the treatment groups show some overlapping regions. On the other hand, it has been documented in the causal inference literature that, if groups differ considerably, then regression models may provide biased estimates of any treatment effects due to extrapolations that can be liable to model misspecification (Li *et al.*, 2013; Drake, 1993).

In order to estimate the neighbourhood effect, we focussed on methods based on a propensity score that allow us to better deal with selection on observables and consequently provide also less biased estimates (Austin, 2011a). Moreover, handling the adjustment for observable confounders separately from the estimation of the treatment effect on the outcome, assumptions about the functional form of the relationship between covariates and outcome are not needed, there are less risks of model misspecification and collinearity among confounders is not a problem (Harding, 2003). However, the use of propensity score techniques in the estimation of neighbourhood effect brings to some methodological complications linked with the fact that we are dealing with neighbourhoods as treatments. The direct consequence is that the number of treatments to be considered is huge, according to the neighbourhood size, and usually applications of propensity score techniques with more than three groups are extremely rare. In our peculiar case, Turin is composed by 10 districts, but it is possible to distinguish also 23 smaller areas and 94 more circumscribed zones.

Main contributions of the thesis

The aim of this thesis is to estimate the neighbourhood effect on old population in Turin with a propensity score approach, in order to do so we need to work on adapting propensity score techniques to work well in a multi-treatments framework with many treatments (starting from ten to higher numbers). Main methodological contribution is divided in two parts: in the first we evaluate the performance of an existing method, usually used in a three treatments framework, in a framework with 10 treatments; in the second part we propose an original algorithm to handle a higher number of treatments.

Applications of propensity score methods in multiple-treatment frameworks remain mainly scattered in the literature, with few applications in three (or four) treatments regimes (Lopez and Gutman, 2017).

In the first part of the thesis, we start from evaluating the effect of the ten districts on subjects' health. Thus, we estimate probabilities of belonging to each neighbourhood (treatment) conditional on observed variables, also known as propensity scores, with Generalized Boosted Models (McCaffrey *et al.*, 2013). The estimated propensity score is transformed to become a weight that recreates a synthetic sample where balance of confounders among treated and untreated subjects is achieved using an inverse probability of treatment weighting approach (IPTW). The Generalized Boosted Models technique, together with other machine learning approaches to compute the propensity score, allows the researchers to fit also complex relations overcoming variables selection and model building processes in an automatic way (Cannas and Arpino, 2018), reducing the risk of misspecification of the model.

In order to evaluate the performance of IPTW in the case of many treatments, we engage in Monte Carlo simulations. We assess IPTW performance under three different scenarios representing different treatment allocations of individuals and compare it with a simple parametric approach, i.e. logistic regression with neighbourhoods as independent variables. In the simulations, IPTW is found to be less biased although it shows a higher variance than logistic regression. This part of the thesis demonstrates that an approach in which treatment assignment and occurrence of the outcome are handled separately is more successful than linear regression models, because it produces less biased estimates. However, increasing the number of treatments using IPTW makes the computational effort to compute propensity scores explode. That is why the application of this approach in a scenario with more than ten neighbourhoods results to be highly inconvenient and nearly impossible.

This last issue is the motivation for the second part of the thesis, where the research goes into the direction to find an useful and practical way to deal with confounders adjustment in a framework with higher number of treatments. In this way, we would be able to estimate the effect of living in one of the 23 areas or 94 zones of Turin. After a review of methods to be applied for balance in multiple treatment frameworks, we ended up with the elaboration of an original proposal that consists on a matching based on Partially Ordered Sets (poset), that we called Matching on Poset based Average Rank for Multiple Treatments (MARMoT). The goal of this method consists in making comparable the characteristics of residents in all the considered areas simultaneously.

Increasing the number of treatments, the computational time to estimate the propensity score grows together with the complexity of the variable that shows the treatment allocation (a categorical one with many levels). In order to overcome this obstacle and summarize confounders' information, we introduced poset theory, that has lately been involved in the construction of synthetic indicators (Boccuzzo and Caperna, 2017). Thanks to poset theory, we are able to assign to each profile, that is each combination of individual characteristics, an approximation of its reciprocal position in the ranking of all profiles. Indeed, if we suppose that individual characteristics that affect living in a specific neighbourhood may be seen as the reflection of a complex latent concept (that includes for instance individuals' socio-economic status and aspirations), it is possible to order individuals' profiles through that concept and match them according to their order. Thus, with MARMoT technique we solved the so called curse of dimensionality, the need to summarize confounders, using the approximation of the rank of these individuals.

Having a score that summarize individuals' characteristics, it is possible to proceed with a matching that assigns to each individual in each neighbourhood one individual in all the other neighbourhoods, and discards those who cannot be matched in order to respect the overlap condition and to make all neighbourhoods comparable. After MARMoT procedure has balanced confounders among neighbourhoods, it is possible to estimate treatment effects using common estimands such as Average Treatment effect on the Treated (ATT).

Before using MARMoT method to estimate neighbourhood effect on real data, we tested it with some simulations with two different scenarios for the allocation of individuals to 23 treatments and two scenarios for the occurrence of the outcome. This technique has proved to be really useful to balance for confounders and reduce bias in estimates.

Thanks to this method, it was possible to estimate the neighbourhood effect on hospitalized fractures in elderly population considering different geographical partitions (10 district, 23 smaller areas and 94 more circumscribed zones) without selection bias due to the different composition of neighbourhoods. These information will be useful for the SCaDU Service to implement prevention policies in the population and urban interventions focusing on neighbourhoods at higher risk.

Chapter 1

Challenges in the estimation of neighbourhood effect

1.1 Introduction

The neighbourhood effect has been defined by Oakes (2004) as the independent causal effect of neighbourhood on a given number of health or social outcomes. Issues linked to the influence of neighbourhood on health and social outcomes have been studied with great interest in the last decades. In this literature review we compare different approaches used to estimate neighbourhood effects listing and discussing some problems and methodological issues. The majority of the analysed studies deals with observational data collected in the United States.

Thus, in literature there are many references to the term neighbourhood, that is often used to delineate person's immediate residential environment and its material and social characteristics that are assumed to have an impact on individuals' outcomes (e.g., deprivation, walk-ability, air pollution, crime and social cohesion) (Diez Roux, 2001). The considered outcomes come from different fields, such as life course events (Rabe and Taylor, 2010), educational achievement (Leckie, 2009) or health outcomes (Cubbin *et al.*, 2000; Pickett and Pearl, 2001). More commonly analysed health outcomes deal with mental health (Kim, 2008; Mair *et al.*, 2008; Truong and Ma, 2006), early child health outcomes (Christian *et al.*, 2015; Sellström and Bremberg, 2006), all-cause mortality (Meijer *et al.*, 2012) and other health outcomes in old population (Roux *et al.*, 2004; Yen *et al.*, 2009). Among principal neighbourhood risk factors there are deprivation, walk-ability, food environment, air pollution, crime and social cohesion (Arcaya *et al.*, 2016). One important concept that these studies disclose is that the exposure to a given neighbourhood with peculiar characteristics can affect health outcomes, this statement reveals to be important in a social epidemiology framework.

Even if the interest in this topic is increasing exponentially, no fully convincing methodological approaches can still be found in literature to estimate neighbourhood effects on different phenomena. Indeed, the main problem for the estimation is that we are dealing with observational data. Thus, the crucial question in most of these studies is whether differences between neighbourhoods can be causally attributed to the context or whether they are simply due to differences between individuals living in different neighbourhoods (Harding, 2003). We call this issue *selection bias*, because the distribution of individuals in the territory (among neighbourhoods) is not at random, it depends on individual characteristics.

Another discussed problem in this field is *endogeneity*, that causes issues both in the estimation and in the interpretation of the neighbourhood effect and that is difficult to explain and recognise because of a lack of agreement on its definition in the literature.

In literature, there are a lot of examples of multilevel models where neighbourhoods are represented by a level. These models are useful to handle the within-group dependence, but they do not overcome the selection bias problem, nor the endogeneity problem (Oakes, 2004). A lot of different studies try to provide a theoretical framework to identify and estimate the neighbourhood effect, among them the most successful is the work of Galster (2008), which is described in the following section.

1.2 Methodological Challenges

A preliminary formulation for the estimation of neighbourhood effects is given by Galster (2008), who pinpointed a systematic list of methodological problems to be discussed in order to try to estimate the neighbourhood effect. In order to quantify neighbourhood effect on an observed outcome (O) at time t for individual i residing in neighbourhood j in metropolitan area k Galster (2008) expressed the following model:

$$O_{ijkt} = \alpha + \beta[P_{ijkt}] + \gamma[P_{ijk}] + \phi[UP_{ijkt}] + \delta[UP_{ijk}] + \theta[N_{jkt}] + \mu[M_{kt}] + \varepsilon$$

where:

P_{ijkt} are the observed personal characteristics that can vary over time, of individual i residing in neighbourhood j in metropolitan area k at time t;

P_{ijk} are the observed personal characteristics that do not vary over time, of individual i residing in neighbourhood j in metropolitan area k;

UP_{ijkt} are the unobserved personal characteristics that can vary over time, of individual i residing in neighbourhood j in metropolitan area k at time t;

UP_{ijk} are the unobserved personal characteristics that do not vary over time, of individual i residing in neighbourhood j in metropolitan area k;

N_{jkt} are the observed characteristics of neighbourhood j in metropolitan area k where individual i resides during time t;

M_{kt} are the observed characteristics of metropolitan area k in which individual i resides during time t.

This basic model raises a list of questions, or challenges, to be faced carefully:

1. The definition of the more appropriate **geographic scale(s)** to conceptualise the neighbourhood N;
2. The identification of the **causal process** that generates the relationship between the neighbourhood and the outcome;
3. The selection of the appropriate characteristics in order to well **operationalise the neighbourhood** N;
4. The identification of the **intensity and the duration** of individual's **exposure** to the neighbourhood N;
5. The selection of **individual's variables** and the definition of strategies to minimise the bias from omitted unobserved individual's characteristics;
6. Identifying and facing the problem of **endogeneity** trying to minimise the bias in the neighbourhood effects estimation.

These are the first six challenges proposed by Galster (2008), but questions to be answered and challenges in order to estimate the neighbourhood effect are still not fully covered. In the following are expressed other interrogatives that can be found in literature, on which a deepen deliberation could drive to better results:

- In Galster (2008) paper only observable time variant neighbourhood characteristics are considered, but **time-invariant and unobservable neighbourhood characteristics** should be taken into account as well, at least in a theoretical representation;

- In the real world the allocation of individuals in the territory is not completely at random, each person consciously chooses and decides where to live according to his/her own characteristics and possibilities. This process creates a **selection bias**;
- The **social stratification** may confound neighbourhood comparisons absorbing all the between-neighbourhoods variability (Oakes, 2004);
- In order to justify **extrapolation**, the exchangeability of individuals between neighbourhoods is usually assumed in a model, which disagrees the real world logic (Oakes, 2004);
- In a more dynamic perspective, the movement of groups of people with similar characteristics between two neighbourhoods may create an effect called **disequilibria** (Oakes, 2004);
- There are no elements to guess that the neighbourhood effect has a linear behaviour on individual's outcomes, thus in literature several authors suppose that it is likely a **threshold or non-linear effect** (Galster, 2014).

As already mentioned, the interest in this work is in investigating the effect that each neighbourhood (treatment) has on individual's health outcomes, adjusting for individual's characteristics (confounders). These relationship may be summed up with the graph in figure 1.1. All the issues mentioned before may be linked with some element in the graph 1.1 and grouped in four thematic areas:

1. **Representation of neighbourhoods:** define the scale and the size of neighbourhoods, identify relevant neighbourhood's characteristics and measure the exposure to neighbourhood (section 1.2.1).
2. **Representation of individuals:** identify relevant individual's characteristics to be included in the analysis (section 1.2.2).
3. **Description of the relation between neighbourhood and outcome:** identify mechanisms of neighbourhood effect and the functional form that better express the neighbourhood effect (section 1.2.3).
4. **Description of the relation between individual's characteristics and neighbourhoods:** two main issues are present, both of them are linked to the fact that we are dealing with a two-way relation. The first issue is that individual's attributes and neighbourhood's characteristics may be mutually causal and it is difficult to recognise

causes and effects (reflection problem), the second problem deals with the fact that individuals chose their own treatment, the neighbourhood in which they live, according to their own characteristics and preferences (selection bias)(section 1.2.4).

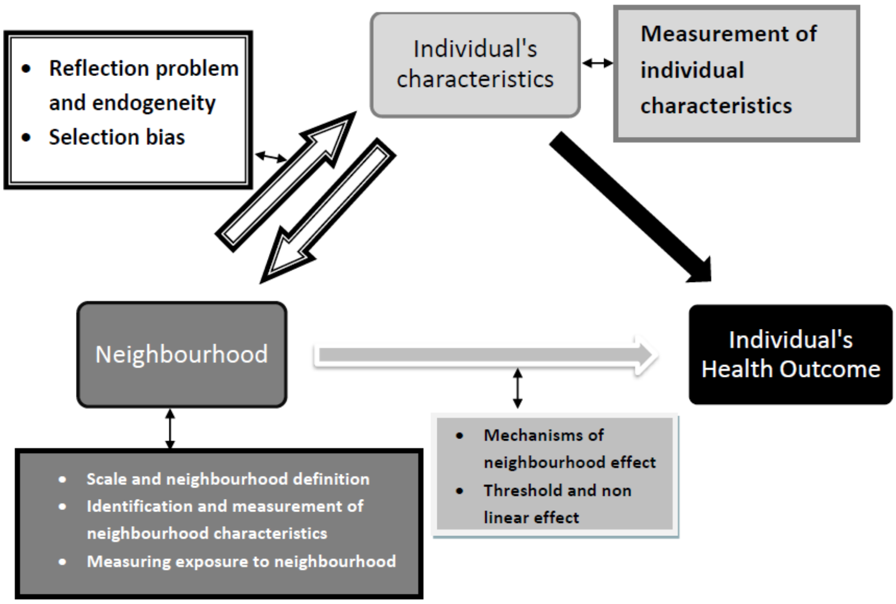

FIGURE 1.1: Summary of methodological problems to be faced in the estimation of neighbourhood effect.

1.2.1 Representation of neighbourhoods

1.2.1.1 Scale and neighbourhood definition

The first issue deals with the definition of the neighbourhood and its scale. There are ambiguous guidelines on how to identify the correct neighbourhood dimension in literature, in fact several possible definitions are available. This brings on many ambiguities because together with the size of the neighbourhood also the neighbourhood effect can change both in intensity and in direction (Galster, 2008). Different criteria to delineate neighbourhood boundaries can be adopted depending on the application: they can be historical, based on people characteristics, on administrative boundaries or based on people's perceptions (Diez Roux, 2001). Those criteria do not necessarily overlap and the choice between them is not trivial, because different neighbourhood definitions lead to different meanings of it and a different effects on individual's outcomes.

According to the review of Galster (2008) about neighbourhood effect studies, there is a strong relation between the definition of the neighbourhood and the local political and institutional boundaries. In U.S., the considered neighbourhood areas are mainly the census tract, which are possibly demographically homogeneous areas containing roughly 4,000 inhabitants, on average. In Western Europe it is possible to find a greater variety of scales: postal codes, "city districts" of various sizes among nations, wards (similar to tracts) and others. For instance, U.K.-based studies used administrative data from wards, lower super output areas (roughly 1,400 inhabitants), and school catchment areas. Moreover, in Italy the census is organized through small geographical partitions that are called census sections, that include on average 170 individuals; starting from them, some aggregations are possible according to the geographical zone.

Usually in these studies the choice of neighbourhoods' boundaries is strictly connected with the available data without a proper attention to its scale and meaning. The importance of the choice of the most meaningful size of the spatial data is taken into account by a serious analytical issue named modifiable areal unit problem (MAUP). In order to include the spatial dimension in quantitative analysis, the areal unit definition is something to deal with carefully, instead of assuming boundaries and areas in the name of pragmatism (Manley, 2014).

On this topic, Manley (2014) described two aspects of the modifiable areal unit problem: the scale effect and the zonation effect. The scale effect consists in having different areal results if we analyse the same phenomenon on several scales; the choice of the scale can drive also to a trade-off between availability of data for a small scale unit system and to a loss of local detail having bigger aggregated areas. In fact, according to Buck (2001) and Bolster *et al.* (2006) research, comparing results on different scales, the neighbourhood effect has higher intensity on smaller scales. In the two last mentioned works, the scale problem is partially solved and confronted by adopting and comparing estimates with different scales and discussing results.

The zonation effect is about how the space has to be divided up; thus, results can vary if, even keeping constant the number of different considered areas, alternative areas aggregations are taken into account. According to Manley (2014), the delineation of areal boundaries is even more arbitrary than the choice of the number of units to consider (which is in other words the scale problem) and can lead to even more unstable statistical analysis.

Another important issue to be considered in the analysis of neighbourhood effect is the Uncertain Geographic Context Problem (UGCoP), which refers to circumstances in which there is a limited knowledge about which geographic scale is causally relevant for health (Arcaya *et al.*, 2016). Thus, the neighbourhood of residence is only one of the places people spend their time, and it might not adequately capture people's exposure to relevant contextual influences. Indeed, according to Kwan (2012), it is important to consider also space in a more dynamical perspective in order to take into consideration that people move around to undertake their daily activities.

Moreover, in the majority of paper about neighbourhood effects estimation, spatial relationships among areas are not considered, nor whether neighbourhoods are situated within larger geographies (Arcaya *et al.*, 2016).

According to Galster (2008), a particular intriguing and promising approach is to consider *bespoke neighbourhoods*, which are concentric circles of varying radii centred on each analysed individual. In this way it is possible to define different radius measures and to try different scales (as in Bolster *et al.* (2006) work); unfortunately, this approach requires a big amount of data with a peculiar territorial detail, which is not so common. A further development of this technique can be represented by overlapping the topographical and street patterns to quantify the radius using a distance not merely geographic, but that may takes into account also the potential for street-level interactions (Grannis, 1998).

1.2.1.2 Identification and measurement of neighbourhood characteristics

The operationalisation of neighbourhood characteristics is not so trivial, it implies the use of some well calibrated proxies in order to measure different categories of neighbourhood effects. Some potential effects may be measured with proxy variables from administrative data, but surveys data with multi-item scales combined in order to retrace social phenomena such as social networks, inter-group interactions and stereotypes, perceptions of disorder and anti-social behaviour, neighbourhood evaluations, etc. are often used in literature (Galster, 2008).

In order to obtain good proxies to describe neighbourhoods, a great effort is necessary and resource-intensive data collection activities are required (Galster, 2008). Neighbourhood characteristics that may be useful to consider are deprivation, walk-ability, air pollution, crime and social cohesion, among others. However, the processing of good composite indicators to sum up neighbourhood characteristics from census data is possible and can even drive to satisfying results.

One of the most successful operationalisation of neighbourhood characteristics is given by Sampson *et al.* (1997), where a *collective efficacy index* was developed as the aggregation of two multi-item scales of *informal social control* and *social cohesion and trust*. The index has been validated using three composite factor-scores indexes based on ten aggregate neighbourhood-level census variables. The three factor were: concentrated disadvantage (composed with poverty, receipt of public assistance, unemployment, female headed-families, density of children and percentage of black residents), immigrant concentration (made of the percentage of Latinos and the percentage of foreign-born people) and residential stability (linear combination of the percentage of people living in the same house as 5 years earlier and the percentage of owner occupied homes). These three factors explained the 70% of the neighbourhood variation in collective efficacy that is defined as social cohesion among neighbours combined with their willingness to intervene on behalf of the common good (Sampson *et al.*, 1997).

1.2.1.3 Measuring the exposure to neighbourhood

Another important point to be explored in order to estimate the neighbourhood effect is the degree to which individuals are exposed to mechanisms that convey neighbourhood effects. Moreover it can be difficult to figure out if those mechanisms have an immediate effect, if their influence has an impact with a temporal lag or if the effect is cumulative through time (Galster, 2008). Anyway the effect is supposed to be stronger for those individuals who have mostly intra-neighbourhood social relationship and for those who have lived in the same place for an extended time period.

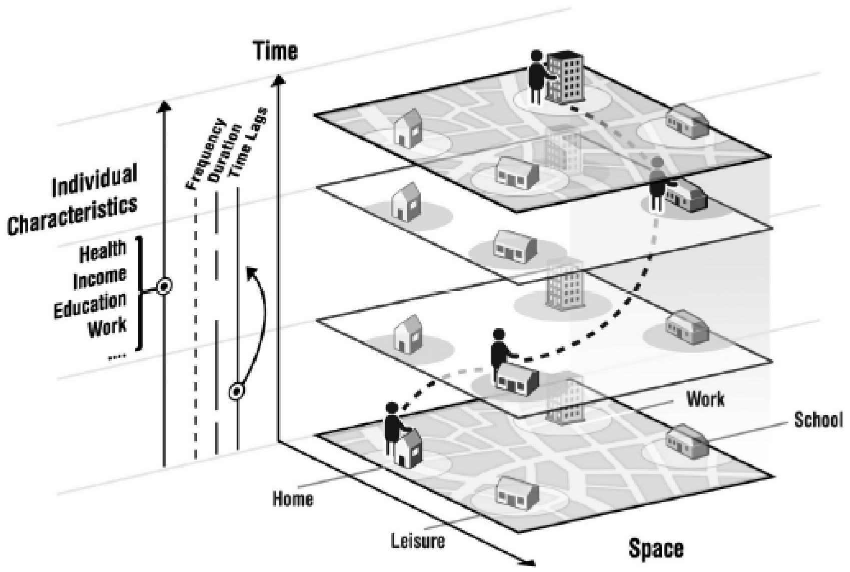

FIGURE 1.2: A conceptual model of the life course approach to neighbourhood effects (de Vuijst *et al.*, 2016).

One recent working paper by de Vuijst *et al.* (2016) proposes a life course approach to study neighbourhood effects, that should represent a comprehensive and dynamic spatial-temporal framework, as shown in Figure 1.2.

Even if the temporal dimension seems fundamental in order to determine the intensity of the neighbourhood effect, the time dimension remains implicit in many studies, receiving still limited attention. Part of the problem deals with the availability of useful data in order to implement more deepened analysis. The conceptual model proposed by de Vuijst *et al.* (2016) is based on the idea that an individual follows a peculiar time-space path over his/her life course, being influenced by various spatial contexts such as the residential neighbourhood, but also by other contexts such as places of work, leisure and schools.

1.2.2 Representation of individuals

According to Galster (2008), individual characteristics that have to be included in the model to estimate the neighbourhood effect can be both time variant and invariant, and they can both observable and unobservable.
Main problems concern the presence of unobservable individual characteristics that may drive to biased estimates of the neighbourhood effects. Thus, even in the most comprehensive dataset, some variables may still remain unobserved; some of these may influence also the selection process that deals with the allocation of individuals among neighbourhoods. Any observed relationship between neighbourhood conditions and outcomes for such individuals may therefore be biased because of this systematic spatial selection process, even if all the observable characteristics are controlled (Galster, 2008).
A possible strategy proposed by Galster (2008) consists in the analysis of only those subjects who did not moved during the observation period. This expedient should be able to remove the selection due to the unobservable variables. However, this peculiar sample selection inserts a new source of selection bias linked to different propensities of individuals to move according to their own characteristics.

1.2.3 Description of the relation between neighbourhood and outcome

1.2.3.1 Mechanisms of neighbourhood effect

Neighbourhood effects are community influences on individual, social or economic outcomes. Examples include health outcomes, labour force activity, child outcomes, criminal behaviour, and other socio-economic phenomena (Dietz, 2002). Neighbourhood is a multi-dimensional package of causal attributes, thus, each part of the package will need to be identified and measured directly (Galster, 2012). Several mechanisms can be found in literature to explain these influences, and to better analyse them in both a qualitative and quantitative way. A possible classification, based on a systematic literature review, has been proposed by Galster (2012):

socio-interactive mechanisms involve social processes endogenous to neighbourhoods (for example social-networks, social cohesion, parental mediation, competition);

environmental mechanisms refer to natural and human-made attributes of the local space that may affect directly the mental and/or physical health of residents without affecting their behaviours (for example the exposure to violence, physical surroundings and toxic exposure);

geographic mechanisms imply aspects of spaces that may affect residents' life courses arising because of the neighbourhood's location in a larger political or economical scale (for example relative to less accessibility to job opportunities, called by Galster (2012) spatial mismatch, or related with public services availability);

institutional mechanisms involve actions by those typically not residing in the given neighbourhood who control important institutional resources located there (for instance stigmatization, which implies reduced opportunities because of public areal stereotypes, local institution resources and local market actors) (Galster, 2012).

The causal relationship between urban structure and health of residents, or health inequalities, is currently strongly supported by a considerable amount of studies and literature reviews (Diez Roux, 2001; Pickett and Pearl, 2001; Truong and Ma, 2006; Yen *et al.*, 2009).

A conceptual model: Spatial Opportunity Structure Model (Galster and Sharkey, 2017) Galster, in his works, tried to find an explanation for the spatial foundation of inequalities. He conceptualised the influence of space on people's socio-economic outcomes with a model that explains the *spatial opportunity structure*. In this model the space works as a *mediator* between personal attributes and achieved status and as a *modifier* of personal attributes effects, as Figure 1.3 shows.

In order to well explain spatial effects on individuals' outcomes it is necessary to start considering the most basic effects: individual's attributes influence on the achieved status (path A in Figure 1.3), the effect that individual's life decisions has on individual's attributes (path C in Figure 1.3) and the influence that parental characteristics have over both individual's attributes and life decisions (path B and D in Figure 1.3). An example to well represent this theoretical model is given by the achieved education level (Individual's achieved status), which is certainly influenced by individual's attitude, such as skills and desire to continue the course of studies. Individual's life decisions directly

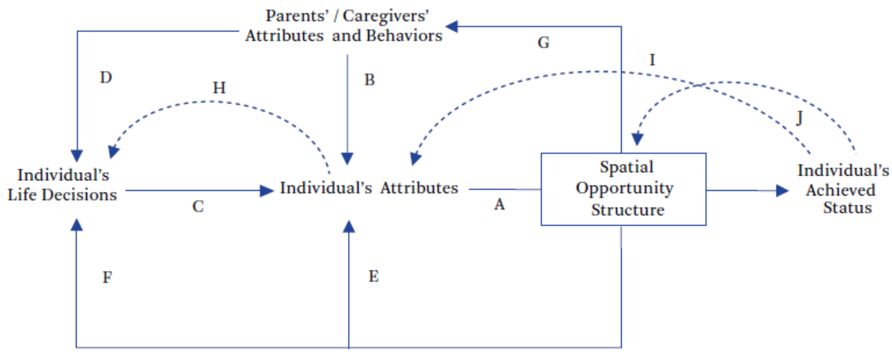

FIGURE 1.3: Conceptual framework of spatial inequalities (Galster and Sharkey, 2017).

influence his/her skills and desires, but they are absolutely conditioned by parents' (or caregivers') attributes and expectations: well cultured parents will expect and incentive their children to reach high educational levels.
Moreover, it is necessary to consider that even the spatial opportunity structure can be a modifier of personal attributes both in a direct and in an indirect way, as we can see from path E, F and G in Figure 1.3. One example of direct effect of the spatial opportunity structure on individual's attributes is given by the school-based peers and collective socialization that can shape individual norms, preferences, aspirations and behaviours, like the school leaving. Some life decisions can also be attributable partially to the spatial opportunity structure, in fact it affects individual's perception of what it is socially more desirable (path F in Figure 1.3). Thus, neighbourhood or school-based peers, role models, and other collective socialization forces can shape a person's norms and preferences, thereby altering the perceived prospective pay-offs associated with various life decisions (Galster and Sharkey, 2017). Moreover, parents and caregivers can alter their parenting styles in response to their perception of the spatial context (Galster and Santiago, 2006), path G in Figure 1.3.

In order to complete the conceptual model, Galster and Sharkey (2017) added three more effects, called feedback effects. The first one, path H in Figure 1.3, deals with the impact that individual's attributes has on Individual's life decisions. Thus, for instance, the acquisition of educational credentials provides a different set of opportunities in individual's life course. The other two feedback effects are caused by the individual's achieved status, which has an effect both on individual's attributes (path I in Figure 1.3) and on the spatial opportunity structure (path J in Figure 1.3). An example of the effect of the achieved status on the spatial opportunity structure consists in the increased income that can guarantee changes in the housing and in the neighbourhood itself.

The conceptual framework, as explained, is quite complex and implies to take care of several methodological and theoretical problems such as endogeneity, stratification, selection bias, reverse causality, intraclass correlation, that are going to be explained in depth in the following pages. Of course, this is only one possible interpretation and theoretical conceptualization of the influence of the neighbourhood on individual's outcomes, but several other models are possible and sustained in literature. Therefore, even in the just explained framework some simplifications are needed in order to estimate and operationalize all the involved effects.

1.2.3.2 Threshold and non-linear effect

In order to well explain this topic, a pharmacological metaphor is useful: the "dosage-response". In this framework it is necessary to look at the neighbourhood as *a multi-element dosage of distal context that generates a response from individual residents that has physical, psychological, attitudinal and/or behavioural manifestations* (Galster, 2014).

Thus, two functional forms can be described:

A non-linear neighborhood effect occurs when the dosage-response relationship is not proportional across all ranges of dosage.

A threshold neighborhood effect is a special case of non-linear effect. It consists in a change of the marginal dosage-response relationship from zero to non-zero (or the opposite) in proximity of a "threshold point".

According to Galster (2014), there are at least three mechanisms through which the neighbourhood has a non-linear behaviour on individual's outcomes. They are socialization, collective social control and stigmatization mechanisms.

The *socialization mechanisms* works as a process of contagion as described by Crane (1991): *if the incidence of problems stays below a critical point, the frequency or prevalence of the problem tends to gravitate toward some relatively low-level equilibrium. But if the incidence surpasses a critical point, the process will spread explosively.*

The *collective social control mechanism* deals with the ability of social norms to modify individuals' behaviour, aspirations and desires. In this case the threshold exists because a group can develop a significant influence in individuals' behaviour only when it reaches some critical mass of density or power.

The *stigmatization mechanism* suggests that when a peculiar group, for instance a "disreputable" one, reaches a critical mass, the public opinion is likely to see the whole neighbourhood as represented by that group and acts consistently. Following the example, if a disreputable group of people gains power and grows over a defined threshold, institutions are inclined to reduce resources and opportunities toward that specific neighbourhood.

Having in mind these mechanisms, it is important to discuss the functional form of the neighbourhood effect without taking the linear form for granted, because it may create bias in the estimates.

1.2.4 Description of the relation between individual's characteristics and neighbourhoods

1.2.4.1 The reflection problem

In order to study and analyse how society affects individual's outcome, a long list of different hypothesis can be taken into account. There is a loss of consensus between economist and sociologists to explain groups dynamics, but even among sociologist several approaches can be found. An interesting interpretation and framework is given by Manski (1995). He explained the identification of different components of society influences on individual's outcomes with a meaningful icon: the reflection of an image in a mirror, calling it the reflection problem. *"Suppose that you observe the almost simultaneous movements of a man and his image in a mirror. Does the mirror image cause the man's movements or reflect them? If you do not understand something of optics and human behaviour, you will not be able to tell."* (Manski, 1995).

In the same way it is difficult to exactly distinguish causes, effects and mechanisms of groups' and individuals' behaviour: is the group behaviour affecting the individual's outcome or is the individual's outcome changing the group composition? Thus, according to Manski (1993), individuals belonging to the same group tend to behave similarly, creating the following effects:

Endogenous effects: the propensity of an individual to behave in some way vary with the prevalence of that behaviour in the group.

Exogenous (contextual) effects: the propensity of an individual to behave in some way vary with the distribution of background characteristics in the group.

Correlated effects: individuals in the same group tend to behave similarly because they face similar institutional environments or have similar individual characteristics.

In order to better understand these three concepts, Manski (1995) considers an example where the outcome is represented by the school achievement of teenage youth. An endogenous effect happens if, *ceteris paribus*, individual achievement tends to vary with the average achievement of the reference group. A contextual effect is observable if the individual achievement tends to vary with the socio-economic composition of the reference group, while a correlated effect is present if students in the same school tend to achieve similarly because of the same environments and backgrounds (families and teachers).

The distinction of these three social effects is more important in a policy implementation framework, because, for instance, endogenous effects are interested by social multiplier phenomena (implementing a policy to improve individual achievement of few individuals, the average of the whole population is enhanced).

Endogeneity. According to Manski (1995), neighbourhood effects are emergent properties of the social interactions of the residents, the behaviour of the neighbourhood influences individual's outcome. Indeed, the endogeneity concept comes with the fact that the neighbourhood is composed of people, so individual characteristics and neighbourhood characteristics may be mutually causal (Galster, 2008). One obvious empirical implication is that some individual characteristics P_{it} and neighbourhood characteristics N_{it} may suffer from multicollinearity (Galster, 2008). Thus, the more an individual is a part of a given neighbourhood, the more his/her characteristics will be similar to neighbourhood characteristics and the neighbourhood effect on the individual's outcome becomes more difficult to identify and to estimate.

Endogeneity is translated in the empirical difficulty to estimate an unbiased neighbourhood effect by means of a linear (multilevel) regression model. Indeed, even after conditioning on some confounders, Z, it is not possible to estimate the conditionally independent effect of X on Y if context effects are endogenous, so by definition not conditionally independent (Oakes, 2004).

1.2.4.2 Selection bias

The problem of selection bias deals with the fact that individuals live in neighbourhoods according to their attributes which are themselves related to the outcome (Diez Roux, 2004). This sort of self-selection causes the social stratification, in fact individuals move through the territory in order to stay close to people with similar characteristics and to live in a suitable neighbourhood. According to these movements, peculiar profiles of individual characteristics may be extremely common in some neighbourhoods and completely absent in some others, and this creates problems in exchangeability hypothesis. For example, neighbourhoods with bad characteristics and less or worse public services will have a high concentration of individuals with low socio-economic status, while these people will not live in neighbourhoods that present better characteristics and likely higher rents.

If the neighbourhood represents the treatment assigned to an individual, then the selection bias issue becomes clearer, indeed in this case the treatment is not assigned exogenously to individuals, thus the correspondent estimated coefficient in a linear regression model is not legitimate to be called *neighbourhood effect* (Manski, 1995).

The self-selection of individuals in their treatment group can be partially controlled using observable confounders, but sometimes variables that are involved in the selection process are not fully observable. There are several different approaches to try to overcome the selection bias problem, based on different hypothesis regarding the set of variables to be considered in the selection process (Galster and Sharkey, 2017):

- Difference models based on longitudinal data can be used if the bias comes from unobserved time invariant variables.
- Fixed-effect models on a siblings sample, but the related estimators have usually big standard errors and they do not control for time-variant unobservable characteristics (Aaronson, 1998).
- Instrumental variables for spatial context characteristics may be devised, but the related estimator presents three different limitations: high standard errors, unrealistic assumptions and it is relevant just for a sub-sample of the analysed population (for those whose assigned neighbourhood was defined by the instrument).
- Propensity score techniques, if it is realistic to assume that individuals matched by observable characteristics are likely to be matched on their unobserved characteristics as well.

Another problem raised by Oakes (2004) is social stratification, that consists on the fact that individuals are stratified among neighbourhoods according to their socio-economical status. This is the direct observable consequence of selection bias issue. The selection equation of any given person observed in a neighbourhood is nearly identical for those who lives in the same neighbourhood and almost completely different for individuals residing in other neighbourhoods, because of the social selection. Having social stratification means that neighbourhood are really different between each others, wandering if the between-neighbourhood differences in the outcome are due to the composition of the neighbourhood or to some exogenous neighbourhood characteristic.

Indeed, in Galster (2012) work, are well described mechanisms that influence individual's outcomes and that can be seen as exogenous or, at least, not involved with the social stratification, such as environmental (as pollution), geographical (as spatial mismatch) and institutional mechanisms.

Moreover, social stratification brings on another important issue for the estimation. Indeed, if the individuals' characteristics distribution among neighbourhoods changes deeply, neighbourhood are not comparable between each other and problems regarding extrapolation need to be faced for the estimation of neighbourhood effect. For instance, in the implementation of a multilevel model, the interest is in the estimation of intercepts and slopes in order to well represent and fit observed data. The main point here is that the theoretical model is made by some straight lines that are extended infinitely in two directions by means of the linear extrapolation (Oakes, 2004). Even if some peculiar cases are not observed, or even impossible to observe, the model assumes a predicted value based on observed data and often biased, because of the exchangeability assumption. Exchangeability is an important assumption that is usually given for granted and it assumes that individuals may be moved from one neighbourhood to another, as if they are exchangeable. However, it may result unsuitable in in the real world. For instance, in figure 1.4, it is represented a multilevel model to estimate the neighbourhood effect on a health outcome, considering three neighbourhoods (B, C, and D) and the socio-economic status (SES) of individuals. The linear extrapolation extends the model infinitely as a straight line, even in areas where the model is not supported by observed data, such as in correspondence of low SES levels even for neighbourood D. In the case shown in figure 1.4 neighbourhoods B and D are not comparable because they include individuals with different levels of SES. Instead of blindly trust exchangeability, it is desirable to check for overlap, especially in situations in which we can expect the presence of social stratification. Indeed, if there is little overlap between neighbourhoods, the estimation will be based mostly on extrapolation and inference will not be supported by data, having an identification error.

Disequilibria In a causal framework, a treatment given to an individual should not have any effect on the assignment of the treatment to other individuals, Rubin (1976) named this important hypothesis the stable unit-treatment assumption (SUTVA). In neighbourhood effect estimation there is the risk of violation the SUTVA. Thus, for instance, moving a large number of poor people to a wealthy neighbourhood reduces the wealth of the target neighbourhood (Oakes, 2004). Characteristics and compositions of neighbourhoods may be modified by people movements across territory, this means

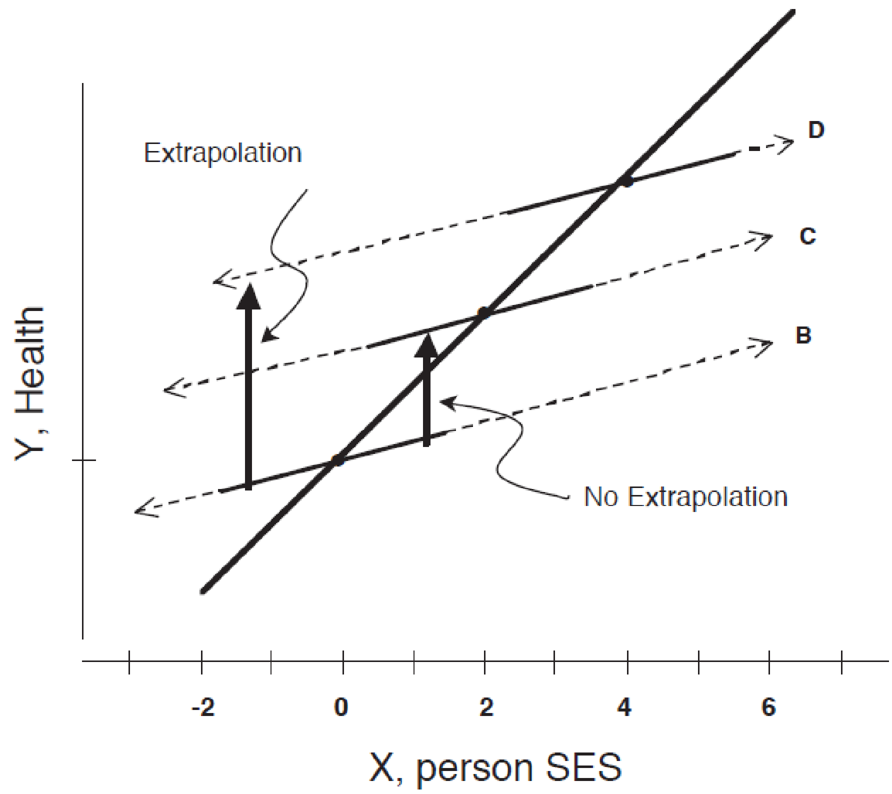

FIGURE 1.4: Extrapolation in arbitrary multilevel models to estimate neighbourhood effects. The graph shows the within and between relationship of SES and some health outcome and highlights the ”within slopes” of three arbitrary neighborhoods, B, C, and D (Oakes, 2004).

that the self-attribution of an individual to a given neighbourhood can affect the context causing changes in the treatment of other individuals already residing in the observed neighbourhood. This aspect has to be taken into account in the analysis in order to select the observed population carefully and avoid disequilibria issues.

1.3 Methodological approaches to estimate neighbourhood effect

1.3.1 Community trials

The randomized community trial is the canonical design for social epidemiology, in particoular for neighbourhood effect studies (Oakes, 2004). They are a valid instrument to measure the neighbourhood effect, but they are expensive, demanding, difficult to perform and unethical. In the 90ties a social experiment had been performed by the Department of Housing and Urban Development (HUD) in United States, it was called

Moving to Opportunity for Fair Housing (MTO). In this experiment a sample of extremely poor subjects has been moved in a different neighbourhood. In that case the main goal was to remove or at least reduce the problem of social selection and segregation in neighbourhoods, but the study design presented some limitations: the sample was not representative of the whole population because only voluntary and poor subjects were involved in the study, causing selection bias (Harding, 2003). Anyway, this social experiment was a fundamental turning point in the estimation of neighbourhood effect, providing robust results regarding effects on mental health and perceived security (Small and Feldman, 2012).

Difficulties in the implementation of experimental studies to develop the neighbourhood effect estimation topic caused the dissemination of observational studies and the consequent development of methodologies to face the new limitations delivered by observational studies.

1.3.2 Multilevel models

The concept of considering the context around analysed subjects represents a fundamental base for social epidemiology and it can be translated in a more theoretical statistical framework as clustering. This is the main reason to use multilevel regression models (Merlo *et al.*, 2005; Caceres *et al.*, 2013). Multilevel studies that examine the relationship between neighbourhood and health had a great growth in the period from 1995 to 2014, with observational data (Arcaya *et al.*, 2016).

Thus, the most common approach to estimate the neighbourhood effect consists in multilevel regression models which allow the simultaneous examination of the within and between neighbourhood variability of the considered outcome. In these models usually neighbourhoods are represented by one level (the highest in the hierarchy) of the model, while other levels usually are the individual (lowest one) and the family one (as for example in Garner and Raudenbush (1991) work).

Since these models include both individual and neighbourhood characteristics, it is possible to evaluate neighbourhood effects on the outcome controlling for individual confounders. The theoretical form of these models is the same proposed by Galster (2008) and cited in section 1.2.3.1, but, some of the multilevel models' assumptions are not perfectly fitted by the empirical framework involved in the neighbourhood effect estimation (Oakes, 2004), creating also some methodological unsolved problems.

Regression models can help adjusting for observed confounders, when the treatment groups present some overlapping regions. However, in the causal inference literature, it has been documented that if groups differ greatly, i.e. social stratification is present, they may provide biased estimates due to extrapolation that can be sensitive to model misspecification (Li *et al.*, 2013; Drake, 1993). In other words, multilevel models just smooth over areas that does not have a common support.

Multilevel models need assumptions about the functional form to represent the relationship between the neighbourhood and the outcome and usually it is hypothesized to be linear. However, as it is explained in the dedicated section (section 1.2.3.2), the relation is usually supposed to be non-linear. If we handle the adjustment for observable confounders separately from the estimation of the treatment effect on the outcome, assumptions about the functional form of the relationship between covariates and outcome are not needed.

With a high number of individual's characteristics, collinearity among confounders may create some misspecification problem using regression models.

Even if multilevel models can easily handle a great number of different neighbourhoods, the whole estimation is based on a model that does not take care of some important problems such as endogeneity, selection bias and comparability of neighbourhoods, extrapolation and non-linearity of the effect.

1.3.3 Propensity score techniques

The gold standard in avoiding selection bias is the use of (quasi) experimental data in which households are randomly assigned to neighbourhoods (Manley and van Ham, 2012). Anyway this is not always possible, so the application of techniques to randomize individuals among neighbourhood retrospectively seems to be a good practise. Indeed, according to Antonakis *et al.* (2010), *if treatment has not been randomly assigned to individuals in groups, if membership to a group is endogenous, or samples are not representative between-groups, estimates must be corrected using the appropriate selection model or other procedures (difference-in-differences, propensity scores).* Several gains exist in using propensity score techniques in order to estimate the neighbourhood effect starting from the fact that selection bias issues are solved taking care of the selection equation and leading back the analysis to a (quasi) experimental situation in the first part of the analysis.

Propensity score, the probability to be treated, may be used to adjust for counfounders in observational studies in many ways, through matching, stratification, covariate adjustment or used as a weight (inverse probability of treatment weighting). In all these cases the main goal is the same: reduce the selection bias, leading back the observational study to an experimental one and making different groups (treatments' groups) comparable (Austin, 2011a).

After the propensity score adjustment, the neighbourhood effect estimate can be computed and this allows more flexibility then, for instance, a multilevel model. No functional form between individual's or neighbourhood's characteristics and the effect is given for granted, so non-linear process can be easily estimated. If the outcome is rare, there are less problems of under representation of cases and the estimating process is more precise and balanced. Moreover, after the adjustment, it is possible and easy to estimate the neighbourhood effect on more than one outcome.

Since the estimation process is not based on a linear model, but is it taken on in two steps, there are also less worries about collinearity and miss-specification of the selection model. Moreover, estimates are not based on extrapolation because using propensity score techniques, it is necessary to verify the overlap, the comparablity of individuals' characteristics in different treatment groups, and just comparable subjects are included in the analysis.

Mainly two aspects remain risky and potential source of bias: endogeneity and disequilibria are not solved. Even if these two problems remains, the estimation of neighbourhood effect through propensity score techniques seems a promising direction to be followed.
In the following chapter, some methods based on the counterfactual approach are described, in particular propensity score techniques.

Chapter 2

Propensity score techniques

2.1 Introduction

Some philosophers of science defined causal effects using the concept of "possible worlds." The actual world is the way things actually are. A possible world is a way things might be. The possible world represents in this definition the counterfactual framework.

The best situation in which it is possible to estimate the causal effect is randomization of treatments, when each treatment is assigned completely randomly to individuals in the sample. Unfortunately, randomization is not always possible, thus, in some situations the researcher can not control the treatment assignment. How is it possible to estimate a causal effect when just observational data are available? Some methods exist in causal inference to try to replicate retrospectively and synthetically a randomization. The main idea is to try to replicate the treatment assignment process in order to identify and control for all the variables that are relevant in that equation. Thus, the number of factors involved in the treatment assignment may be huge; in order to overcome this *dimensionality problem*, Rosenbaum and Rubin (1983) proposed the use of a balancing score based on the probability of receiving the treatment given pretreatment characteristics, the *propensity score*.

Propensity score may be involved in the balancing procedures mainly in four different ways (Austin, 2011a):

Propensity Score Matching (PSM), where treated and untreated individuals with the same propensity score values are matched and their outcomes are compared;

Propensity Score Stratification, according to which the propensity score is used to stratify subjects into mutually exclusive subsets where the balance of counfounders should be reached;

Propensity Score Covariate adjustment where the outcome is regressed using the computed propensity score and the assigned treatment; the estimation of treatment effect is given by its estimated coefficient.

Inverse probability of treatment weighting (IPTW) that involves the propensity score as a weight that should recreate a synthetic sample where balance of counfounders among treated and untreated subjects is achieved.

Since the aim of this thesis is to compute the neighbourhood effect, using propensity score matching techniques, in this chapter a brief introduction to these methods is provided, drawing some connections to the empirical application. After an introduction about basics of the counterfactual approach in a binary treatment framework, the four propensity score methods just mentioned are briefly described. Last part of this chapter is dedicated to the extension of the counterfactual approach in a multi-treatment framework.

2.2 Counterfactual approach and propensity score

Let suppose there is a population composed of N individuals, each of them indexed by $i = 1, \dots, N$. Two fundamental variables are associated with each subject: a binary variable T that represents the dichotomous treatment and assumes value 1 if the subject receives treatment 1 (lives in neighbourhood 1) and value 0 if the subject receives treatment 0; and the outcome variable Y. Moreover, each subject has a pair of potential outcomes: Y_{0i} and Y_{1i}, are the outcomes under the treatment $T = 0$ and the treatment $T = 1$, respectively. However, each subject receives only one of the treatments (control treatment or active treatment) (Austin, 2011a). The causal effect for subject i of living in neighbourhood 1 is $\tau_i = Y_{1i} - Y_{0i}$; i.e., the difference between the outcome of individual i who lives in neighbourhood 1 and the outcome for the very same individual if he/she lived in neighbourhood 0, (Holland *et al.*, 1985) and the effect may differ among subjects.

In practice, it is not possible to observe the very same individual living in the same period in two different neighbourhoods. Thus, we can observe just one potential outcome corresponding to the assigned treatment for each individual.

There are two most used causal estimands, that are defined as follows:

- **the average treatment effect, ATE,** is defined as

$$ATE = E(Y_1 - Y_0)$$

that corresponds to the causal estimand for the whole population.

- **the average treatment effect on the treated, ATT,** which is

$$ATT = E(Y_1 - Y_0|T = 1)$$

that is the average treatment effect on those subjects who ultimately received the treatment (Austin, 2011a).

In order to apply propensity score techniques, some assumptions are needed, such as *temporality*, which implies that the treatment selection T must occur before the outcome; the *strong ignorability*, which is composed of two assumptions, unconfoundedness and positivit; and the *stable unit treatment value assumption (SUTVA)*. According to the unconfoundedness assumption, the potential outcomes (Y_1, Y_0) are independent of the treatment (T) assignment, given a set of observable variables X that are not affected by the treatment, $Y_1, Y_0 \perp\!\!\!\perp T|X$. This assumption is also known as selection on observables, as it is based on the premise that there are no unmeasured confounders, because all of the variables that are involved in the selection process have been observed, measured, and included in the propensity score computation. The positivity or overlap assumption is based on the premise that each subject must have a positive probability to be included in the treatment or control group, $0 < P(T_i = 1|X_i) < 1$. The SUTVA includes two assumptions: the *no interference* and the *stable treatment assumption*. According to the SUTVA, the potential outcomes for any given unit do not vary with the treatments assigned to other units; and, for each unit, there are no different forms or versions of each treatment level that lead to different potential outcomes (Imbens and Rubin, 2015).

To be able to include all of the observable confounders, we may have to deal with a large number of covariates; this problem is called the curse of dimensionality, and can be solved with the use of a so-called balancing score (Caliendo and Kopeining, 2008). A balancing score, $b(X)$, is a function of the observed covariates X such that the conditional distribution of X given $b(X)$ is the same for treated $(T = 1)$ and control $(T = 0)$ units; that is, $X \perp\!\!\!\perp T|b(X)$ (Rosenbaum and Rubin, 1983). Rosenbaum and Rubin (1983) have demonstrated that the propensity score e_i, or the probability that each individual has to receive the treatment, $e_i = P(T_i = 1|X_i)$, is the coarsest balancing score. Propensity scores are usually estimated using parametric models, such as a logistic regression. If these models are misspecified, the balance of the covariates may be not satisfactory. This is the reason why several different methods for estimating the propensity score have recently been implemented and compared (Setoguchi *et al.*, 2008; Li *et al.*, 2013), such as some CART-based (Lee *et al.*, 2010) (pruned, bagged, and boosted (McCaffrey *et al.*, 2004)), neural network, and random forests. The use of data mining techniques in this field has been shown to produce more balance and less bias in the causal estimators based on the propensity scores. Indeed, these flexible data-driven algorithms allow researchers to fit complex relations while automatically addressing challenges arising from variable selection and model building processes(Cannas and Arpino, 2018). If the relationship between the counfounders and the treatments is not linear or is not additive, machine learning techniques are able to gather and handle them automatically in the estimation process. Even if these techniques provide models that are difficult to interpret, they represent an important resource for estimating the propensity score, because interpretation is not a fundamental consideration in this step of the analysis, whereas the balance that can be reached with propensity score adjustments is of primary interest.

In order to measure the balance of each confounder X between treatment groups, the Absolute Standardized Bias (ABS) measure, is usually employed:

$$ASB = \frac{|\bar{X}_0 - \bar{X}_1|}{\sqrt{\frac{S_0^2}{2} + \frac{S_1^2}{2}}} \tag{2.1}$$

where $\bar{X}_0$ and $\bar{X}_1$ are the means of variable X of individuals living respectively in neighbourhoods $T = 0$ and $T = 1$; S_0 and S_1 are the standard deviations of variable X for individuals living in neighbourhoods $T = 0$ and $T = 1$, respectively.

2.3 Propensity score matching

The propensity score matching technique consists in forming matched sets composed by individuals of different treatment groups with similar propensity score value (Rosenbaum and Rubin, 1983). This technique is more commonly applied in the one-to-one matching, in which pairs are formed matching one treated individual with a control subject (Austin, 2011a). The aim is to get, after matching, a balanced sample where confounders have the same distribution in the treatment and the control groups. There are different methods to select individuals to form the matched sets to be included in the balanced sample, several choices to define the matching algorithm that better fits the analysed context.

The first one is to choose between matching with and without replacement. For each treated individual it is necessary to find an untreated subject with similar propensity score, when the couple is formed the difference between with and without replacement matching becomes relevant. If the control unit already included in a matched pair is still available to form other couples, then the matching is called with replacement, while if the already matched control unit is no longer available as potential match for other treated subjects, it is a without replacement matching.

Another possible choice is between greedy and optimal matching (Rosenbaum, 2002). In a greedy matching algorithm, the control subject with the closest propensity score is selected to be matched with each randomly selected treated unit, even if the selected control unit would have been the best match for another treated subject. On contrasts, the optimal matching forms couples in order to minimize the total within-pair difference of the propensity score.

The third decision to make involves the choice of the control unit with the "closest" propensity score. There are many matching methods, the two most commonly used are: the nearest neighbourhood and the nearest neighbourhood within a specified caliper distance matching. According to the nearest neighbourhood matching method, given a treated unit, the control subject with the closest propensity score is selected; if there are more than one untreated individuals equally close to the treated one, the second member of the couple is randomly selected among them. However, in this way there are not guarantees about the distance of the propensity scores between the two matched individuals. The nearest neighbourhood within a specified caliper introduce some restrictions that regard the maximum distance between the propensity scores of the subjects included in the matched couple. Indeed, the maximum allowed distance (called also caliper) in this case is fixed. If it is not possible to find a control unit with a propensity score in the range of values defined by the caliper, the treated subject will not be matched with any control unit and it will be excluded from the analysis.

Having in mind to perform a nearest neighbourhood within a specified caliper distance matching, the choice of the dimension of the caliper becomes another important decision to make. In the literature this value is usually proportional to the standard deviation of the propensity score, such as the standard deviation divided by 4 (Cochran and Rubin, 1973; Lunt, 2013) or 5 (Austin, 2011b), and there are some simulation studies that show that a caliper value equal to the standard deviation divided by 5 provides a reduction of approximately 99% of the bias due to the measured confounders (Gu and Rosenbaum, 1993).

After matching is performed, the treatment effect may be computed just comparing the outcome in treatment and control groups in the matched sample.

Propensity score matching is an extensively used method to balance for confounders and in literature it is possible to find several variations of those just described. Moreover, other matching procedures to balance for confounders exist that are not based on propensity score, for instance the template matching (Silber *et al.*, 2014) and the coarsened exact matching (Iacus *et al.*, 2012).

2.4 Inverse probability of treatment weighting (IPTW)

According to the inverse propensity score weighting procedure, the main goal is to create an artificial sample in which the distribution of covariates is independent of the treatment assignment, such that the treated individuals have the same characteristics as the untreated individuals. The balanced sample is composed of the real sample including all observed individuals, but weighted according to the propensity score value. In this case, the weights work as sampling weights (Horvitz and Thompson, 1952; Morgan and Todd, 2008).

Thus, the weights can be defined as

$$w_i = \frac{T_i}{e_i} + \frac{(1 - T_i)}{1 - e_i}. \tag{2.2}$$

In other words, a subject's weight is equal to the inverse of the probability of living in the neighbourhood in which the individual is actually living (Austin, 2011a). With these weights it is possible to estimate the Average Treatment Effect (ATE), using the following expression(Austin, 2011a):

$$A\hat{T}E = \frac{1}{N}\sum_{i=1}^{N} \frac{T_i Y_i}{e_i} - \frac{1}{N}\sum_{i=1}^{N} \frac{(1 - T_i)Y_i}{1 - e_i}. \tag{2.3}$$

The resulting estimator can be seen as the difference between two weighted means. Some individuals may display very low probabilities of being in a given neighbourhood; i.e., high weights. Particularly high weights cancause an inflated variance of estimates, and can also cause a distortion of the results, which is why the trimming of extreme weights has been studied in the literature (Lee *et al.*, 2011). Indeed, this method may have issues when the positivity assumption nearly fails. This method presents some important pros: it requires just a good specification of the treatment assignment model that does not fear the overfitting, including predictors of the outcome, for this peculiar approach because it may be rewarding in terms of precision. Moreover, this method can be easily extend to multi-treatment and continuous treatment frameworks.

2.5 Propensity Score Stratification

The propensity score stratification consists in dividing analysed population into strata according to the propensity score and computing an effect in each stratum. Indeed, the first step is to order individuals according to their propensity score, then strata may be defined using percentiles of the propensity score as thresholds. In literature, five strata defined by quartiles have been used and shown to eliminate approximately 90% of the bias due to measured confounders (Rosenbaum and Rubin, 1984). An increase of the number of strata should improve the bias reduction, but the marginal drop of the bias for each additional stratum will decrease increasing the number of strata.

Once strata are defined, it is possible to estimate the treatment effect within each stratum and the overall effect will be the weighted mean of each stratum-specific treatment effect, using strata's sizes as weights.

Even if this method is easy to implement and gives stable estimators even when the positivity assumption nearly fails, it requires a correct model for the propensity score and fails to remove confonding bias if the within stratum distribution of the propensity score differs between treated and controls individuals.

2.6 Propensity Score Covariate adjustment

In the propensity score covariate adjustment approach, the propensity score is used as an independent variable included in a model to explain the outcome together with an indicator variable denoting the treatment status. According to the nature of the outcome, a linear regression model (for a continuous outcome) or a logistic regression model (for a dichotomous outcome) is fitted and the effect of the treatment is represented by the treatment estimated coefficient and interpreted as an adjusted difference in means (for a continuous outcome) or as an odds ratio (for a dichotomous outcome).

Even if this method seems quite easy to implement, an important drawback is that it requires an assumption about the functional form of the relationship between the propensity score and the outcome.

2.7 Propensity score techniques in a multi-treatment framework

In a multi-treatment framework, it is possible to represent, for simplicity, the treatment assignment with a set of dummies $D_{it}(T_i)$ (Linden *et al.*, 2016), where T_i is a multivalued treatment variable that takes values from 1 to K (in our specific application, it takes values from 1 to 10):

$$D_{it}(T_i) = \begin{cases} 1 & if \quad T_i = t \\ 0 & otherwise. \end{cases} \quad for \quad t = 1, \ldots, K \tag{2.4}$$

Consequently, we will have a set of potential outcomes $\mathbf{Y} = (Y_{1i}, ..., Y_{Ki})$ for individual i considering all different treatments, and just one of them is observed.

In a multi-treatment framework the choice of the most adequate estimand to represent the effect of the treatment is not trivial. Indeed, even considering the already mentioned ATE and ATT, different choices are possible depending on the comparisons the researcher considers of interest. One possibility is to compute the ATE and the ATT as if each neighbourhood is considered as a dichotomous treatment, so the comparison becomes between neighbourhood t and the rest of the city t^c. The specification for the ATE for each treatment t will be

$$ATE_{t,t^c} = E[Y_t - Y_{t^c}], \tag{2.5}$$

while the specification for the ATT for each treatment t will be

$$ATT_{t,t^c} = E[Y_t - Y_{t^c} | T = t]. \tag{2.6}$$

However, in other cases the most informative comparison may be between two neighbourhoods or each neighbourhood and a common reference (for instance the neighbourhood with lowest rate of hospitalized fractures). It is possible to obtain those estimands with just small variations of these formulas (4.3,4.4).

In the multi-treatment case, the same assumptions as in the dichotomous treatment framework are needed, but under the circumstance that there are K counterfactual outcomes, and not just two (Lopez and Gutman, 2017). The SUTVA needs to be extended to a vector of potential outcomes (Imai and Van Dyk, 2004), while the strong ignorability for the multi-treatment framework becomes

- $Pr[\mathbf{Y}|T = t, x] = Pr[\mathbf{Y}|x]$, referring to the unconfoundedness assumption; and

- $0 < Pr[T = t|x] \forall t \in T$, referring to the positivity assumption.

Finally, also for the measure of balance or ASB it is possible to find in literature more than one possible expression according to treatment comparisons of interest. In this work we define the ASB as

$$ASB = \frac{|\bar{X}_t - \bar{X}|}{\sqrt{\frac{S_t^2}{2} + \frac{S^2}{2}}} \tag{2.7}$$

where $\bar{X}$ and $\bar{X}_t$ are the means of variable X of individuals living respectively in the whole population and in the neighbourhoods t; S and S_t are the standard deviations of variable X with respect to individuals living respectively in the whole population and in the neighbourhood t.

In a multi-treatment framework, the propensity score also needs a different specification. Imbens (2000) proposed a modified definition of the propensity score. The generalized propensity score (GPS) is the conditional probability of receiving a particular level of the treatment, given the pretreatment variables. Generalized propensity score applications remain largely scattered in the literature, however, with few applications in regimes involving three (or four) treatments (Lopez and Gutman, 2017). Some of these involve binomial comparisons (Lechner, 2001, 2002) that may pose problems in terms of common overlap and computational effort when the number of treatments increases. Other attempts have focused on forming triplets to compare subjects in a three-treatment framework using matching algorithms (Hade, 2012; Rassen *et al.*, 2011), or larger numbers with vector matching (Lopez and Gutman, 2017). The application of IPTW approaches has been explored by combining different techniques (McCaffrey *et al.*, 2013; Linden and Yarnold, 2016). Other methods that have been tested and compared (Linden *et al.*, 2016) include: regression adjustment (Spreeuwenberg *et al.*, 2010); marginal mean weighting through stratification (Hong, 2010, 2012); and doubly robust methods like the Inverse Probability of Treatment Weighting (IPTW) regression adjustment (Uysal, 2015).

None of these methods are practical, however, if the number of treatments greatly increases. Some important assumptions (such as the overlap) become difficult to satisfy, and estimating the propensity score becomes computationally demanding. The most common model for estimating a GPS is the multinomial logistic regression (Lopez and Gutman, 2017): using this model, K propensity scores e_{it} with $t = 1, ...K$ are estimated, one for each treatment, and they sum to 1. The dependent variable of such a model in a framework with many treatments is therefore categorical with many levels. The result of such a model in a multi-treatment framework would be an estimation of many small probabilities, with small differences between them (generally speaking, with 23 treatments we would expect a mean of the predicted values of around 0.04 for each individual). An alternative approach, to solve the curse of dimensionality without needing to estimate the probability of receiving each treatment, is template matching. This method can handle the balance of many treatments, and it has been used to compare the performance of hospitals, for instance, reducing the bias due to their different case-mix of patients (Silber *et al.*, 2014). Taking this approach, a sample of individuals represented in all the treatment groups is selected so as to make the individuals in all the treatment groups included in the analysis comparable. This sample becomes the template. Then the matching algorithm matches individuals from all treatment groups with the template, and all other individuals are discarded. The analysis is thus restricted to individuals belonging to the common support of covariates across all the treatment groups. The matching procedure remains similar to the binary case, focusing only on the template and its selected variables. The final dataset will comprise a sample of individuals for each treatment group that resembles the template as much as possible. This simplification enables a huge number of treatments to be managed, but limits the analysis to the individuals comprising the template, and to the choice of template. This means that the target population experiencing the estimated effects may differ considerably from the whole sample population, even though it will be relevant with respect to the chosen template.

Chapter 3

Data

3.1 Introduction

The topic of this thesis comes from the real need of the Regional Unit of Epidemiology and Health Promotion in the Piedmont Region about measuring the impact of the Turin neighbourhoods on health among old individuals. Indeed, the thesis is one of the results of a fruitful collaboration with the Unit "SCaDU Servizio Sovrazonale di Epidemiologia" in Grugliasco (Turin, Italy), that is involved with the Department of Statistical Science of the University of Padova in a formal agreement.

ASL TO3 is a leader health unit in Italy and it is involved in several European projects (2015-2018), all concerning health and health inequalities in urban contexts:

- LIFEPATH – Lifecourse biological pathways underlying social differences in healthy ageing;
- HORIZON 2020 (http://www.lifepathproject.eu);
- MINDMAP, Promoting Mental Wellbeing In The Ageing Urban Population: Determinants, Policies and Interventions In European Cities;
- EURO-HEALTHY, Shaping European Policies to Promote Health Equity (http://www.euro-healthy.eu).

The data used in this thesis are part of the health administrative datasets of Turin, Italy, and they have been managed and prepared for the analysis in Turin with the help and under the supervision of Prof. Giuseppe Costa and his collaborators of the SCaDU Service.

In the first part of this chapter the Turin Longitudinal Study is described, in the second section are listed and described all the confounder variables that we used in the empirical application and in the simulations and the third part is about the geographical partitions of Turin.

3.2 Turin Longitudinal Study

The data used in our analysis come from a longitudinal study conducted in Turin, that gave rise to an integrated database, which combines administrative data flows on residents drawn from censuses with health data flows (hospital discharge records, prescription charges and exemptions, and territorial drug prescriptions). The hospital discharge records contain information on the patient's diagnosis, admission modality (emergency, compulsory, voluntary), and dates of admission and discharge. The prescription charges database lists all exemptions from payment of health services to which some patients are entitled due to chronic conditions or low income. The territorial drug prescriptions database contains details of prescribed drugs, the quantities involved, and their classification (based on their therapeutic, pharmacological and chemical properties). The census data includes not only basic demographic details, such as age, sex, and place of birth, but also some important information about individuals' socio-economic status, such as their occupation, education, home ownership, and family composition.

All these different data sources have been pooled together over time (see figure 3.1). Starting with the censuses and population registries available in 1971, Turin's residents have been registered and tracked as a historical migration dataset, considering all movements of individuals living in Turin for at least one day from 1971 onwards (Costa *et al.*, 2017). Several other data sources were added over time, such as the cause of death archives in 1971, the cancer registry in 1985, the hospital discharge records in 1995, drug prescriptions data in 1997, and so on. In this work we are dealing with big-data, since we have a lot of information about the whole population of the city of Turin (table 3.1). Thus, the population of Turin is around 850.000 and dimension of archives that are part of the Longitudinal Study in Turin are huge: for instance, the hospital discharge records counts 3.219.996 records from 1995 to 2013 and the territorial drug prescriptions database contains 119.146.470 records from 1997 to 2013.

In order to get the variables of interest and to select the population for the analysis, several deterministic record linkage based on anonymous identification codes were performed.

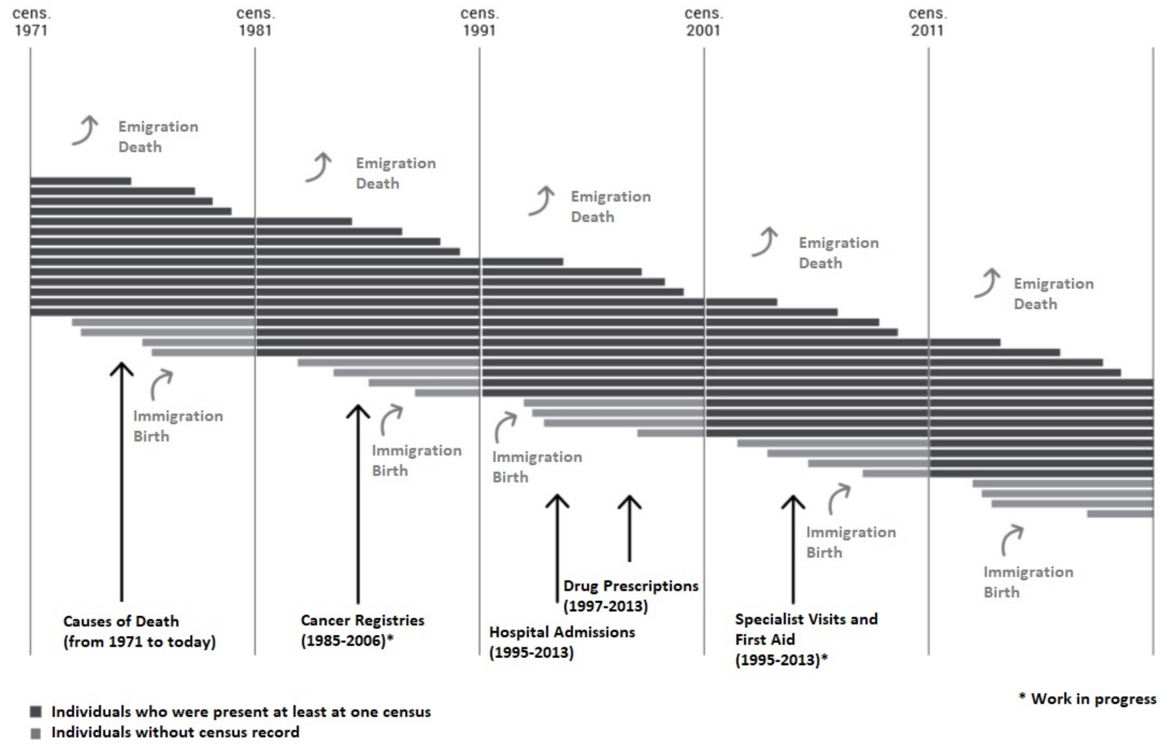

FIGURE 3.1: The Turin Longitudinal Study, 1971-2013. Source: (Costa *et al.*, 2017)

TABLE 3.1: Data sources included in SLT (Costa *et al.*, 2017).

Source	**Records**
Civil Registry (at 30/08/2015)	2391833
Residential History (1971-2015)	4566180
Parenthood linkage (1971-2011)	996806
Census 1971	1023578
Census 1981	1091287
Census 1991	930072
Census 2011	849686
Causes of Death (1971-2013)	490869
Cancer registries(1985-2006)	123078
Hospital admissions (1995-2013)	3219996
Drug prescriptions (1997-2013)	119146470
Specialist visits and First Aid (2002-2005)	44258010

3.2.1 Variables

Based on the literature on neighbourhood effects on older people's health (Roux *et al.*, 2004; Yen *et al.*, 2009), we consider the following variables as possible confounders: gender, age (considering five-year age brackets: 60-64, 65-69, 70-74, 75-79, 80 and over), region of birth, marital status, family composition, educational attainment, last known occupational condition, ethnicity (that distinguish between individuals born in Italy or outside of Italy), living alone, type of housing, overcrowding and home ownership. The region of birth is coded, distinguishing between individuals born: in Piedmont (the region to which Turin belongs); in other regions of northern Italy; central Italy; southern Italy or islands; or outside Italy. Marital situation distinguishes between never married, married, widow or divorced. The variable representing family composition combines marital status with the number of components: living alone; married and living only with partner (family of two); unmarried and not living alone (family of two or more); married and living in a family of more than two people. The last known occupational situation is a variable obtained from the census data from 1971 to 2001, and aims to capture the last type of occupation prior to retirement. This was not possible for some individuals because they were already retired in 1971 (or in all the censuses concerning them), or they were not working for other reasons. The occupation variable distinguishes between the above-mentioned case and home-makers, entrepreneurs, white-collar workers, and manual workers. The variable living alone has been composed using previous censuses, it distinguishes between individuals that do not live alone at 2001 census date, individuals that live alone at 2001 census date but not at 1991 census date and individuals that were already living alone at 1991 census date. Type of housing is a variable that describes the house in which each individual live and it is an ordinal variable with three levels that describe the wealth and comfort of the house. The variable overcrowding consists on the ratio between the number of rooms and the number of family components.

Moreover, we consider variables that describe health conditions of individuals in neighbourhoods: diagnoses of hypertension or cardiac issues and the number of different kinds of drugs that have been prescribed to individuals. We do not use these variables in the empirical study because they can be themselves be affected by the treatment. However, they are included in the sample of variables that are considered in the simulation study because in that context we are manipulating the true data-generating models.

3.3 Neighbourhoods in Turin

The city of Turin can be split into 10 districts, 23 areas, or 94 zones, considering neighbourhoods that might be causally relevant to health (Arcaya *et al.*, 2016). The three partitions may relate to different living conditions (deprivation, walkability, crime, and social cohesion) and population characteristics, but the three geographical layers are only partially hierarchical. For instance, the same zone may belong to two or more areas, or districts.

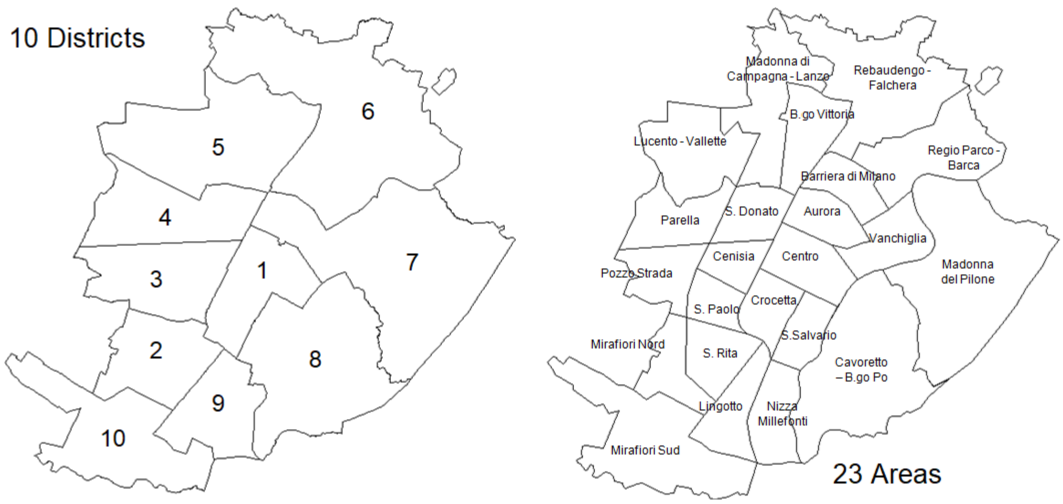

Figure 3.2: Geographical partition of Turin in 10 districts and 23 areas.

Chapter 4

Evaluating inverse propensity score weighting in the presence of many treatments.

4.1 Introduction

In this chapter, we propose using inverse probability of treatment weighting (IPTW) to deal with the non random allocation of individuals to neighbourhoods, thus making neighbourhoods comparable with respect to observable individuals' characteristics. IPTW has been previously assessed with a limited number of treatments (2 or 3). However, in estimating neighbourhood effects a considerably higher number of treatments (i.e., neighbourhoods) arises. In this chapter we engage in Monte Carlo simulations to evaluate the performance of IPTW in the case of many treatments. Specifically, using Generalized Boosted Models, we estimate probabilities of belonging to each neighbourhood (treatments) conditionally on observed variables, also known as propensity scores. We assess IPTW performance under three different scenarios representing different treatment allocations of individuals and compare it with a simple parametric approach, i.e. logistic regression with neighbourhoods as main independent variables. In the simulations, IPTW was found to be less biased although it showed a higher variance than logistic regression.

In the first part of the chapter the IPTW approach is described for the multi-treatment framework, focusing on the approach proposed by McCaffrey *et al.* (2013). Then the simulation study is described with details on the design and results.

4.2 IPTW in multi-treatment framework

The IPTW estimator based on propensity score presented in section 2.4 for the dichotomous case can also be extended to the multi-treatment framework, in which the propensity score needs a different specification.

The most common model for estimating a GPS is the multinomial logistic regression(Lopez and Gutman, 2017), which produces K propensity scores e_{it} with $t = 1, ...K$, one for each treatment, that sum to 1. Having the propensity score e_{it}, weights are defined as

$$w_i = \sum_{t=1}^{K} \frac{D_{it}(T_i)}{e_{it}}. \tag{4.1}$$

In a multi-treatment framework, the identification of the most adequate estimand to represent the effect of the treatment is not trivial. Indeed, even considering the already mentioned ATE, it is possible to compute this estimator in different ways depending on the purpose of the study, while assuming different comparisons. For instance, if the aim is to evaluate the causal effect of living in a given neighbourhood t' versus another specific neighbourhood t'', the ATE may be computed through the following formula (Linden *et al.*, 2016):

$$A\hat{T}E_{t',t''} = \frac{1}{N}\sum_{i=1}^{N}\frac{Y_i D_{it'}(T_i)}{e_{it'}} - \frac{1}{N}\sum_{i=1}^{N}\frac{Y_i D_{it''}(T_i)}{e_{it''}}. \tag{4.2}$$

Alternatively, it may be more informative to compare each neighbourhood t with all of the others, considering all possible couples and computing $\binom{K}{2}$ pairwise ATE. Otherwise, the comparison with the rest of neighbourhoods may be considered the more explicative for the analysed phenomenon. It is possible to obtain those estimators with just small variations of formula 4.2.

As anticipated in the introduction, parametric models are often used in neighbourhood observational studies. For example, it is possible to use a logistic regression with dummy variables representing $K - 1$ neighbourhoods and controlling for confounders. To make the comparison between this approach and an approach based on IPTW easier, we estimated a logistic regression weighted by the inverse probability of treatment, as we would do with sampling weights. More specifically, we consider odds ratios, $OR = e^{\hat{\beta}}$, as the estimand of interest.

In order to implement the IPTW approach, we adopted an approach proposed by McCaffrey *et al.* (2013) for a multi-treatment framework. This method is based on a Generalised Boosted Model (GBM) for computing the propensity score while reducing the risk of the misspecification of the treatment assignment model; and it is implemented in the `twang` package in R (Ridgeway *et al.*, 2006) (Toolkit for Weighting and Analysis of Nonequivalent Groups).

The first part of the algorithm consists of estimating the propensity score as in a dichotomous framework, while considering the treatment groups separately and balancing them with the whole population. For each neighbourhood t, the GBM fits a piecewise constant model composed of many simple regression trees in order to predict the dichotomous treatment "living in neighbourhood t or elsewhere in the city". These regression trees are combined to iteratively adjust the log-odds of treatment assignment $g(\mathbf{X})$ in order to maximise the log-likelihood function:

$$\ell(g) = \sum_{i=1}^{N} D_{it}(T_i)g(\mathbf{X}_i) - \sum_{i=1}^{N} log\{1 + exp[g(\mathbf{X}_i)]\} \tag{4.3}$$

where $D_{it}(T_i)$ is the treatment assignment indicator and $\mathbf{X}$ contains all the confounders (McCaffrey *et al.*, 2004). The iteration proceeds until the stopping rule is satisfied; in this case, it regards the balance of pre-treatments covariates. A possible balance measure is the Population Standardized Bias (PSB) for each variable v and each neighbourhood t'. This measure compares the distribution of the confounders in each treatment group and in the whole population, while considering all treatment groups. It is given by the formula:

$$PSB_{vt'} = \frac{|\hat{\bar{X}}_{vt'} - \hat{\bar{X}}_{vp}|}{\hat{\sigma}_{vp}} * 100, \tag{4.4}$$

where $\bar{X}_{vt'}$ is the mean of variable v computed on the analysed sample weighted with the inverse of the propensity score of being in neighbourhood t', and $\bar{X}_{vp}$ and $\hat{\sigma}_{vp}$ are the unweighted mean and the standard deviation of variable v in the whole population (McCaffrey *et al.*, 2013). According to the literature, the possible thresholds for defining a balanced population are 25%, 20%, and 10% (Austin, 2009; Rosenbaum and Rubin, 1985).

The PSB balance measure is computed automatically for each variable and each neighbourhood, and needs to be summarised. It is possible to chose between two summary statistics: namely, their mean value or their maximum value among all considered covariates.

As we have a lot of dichotomous variables in our analysis, we have decided to use the Population Standardized Bias to measure the balance among the covariates; and, to be more conservative, to summarise it by its maximum value (instead of using the mean) among the pre-treatment variables. Indeed, minimising the maximum PSBs guarantees that all of the other values are smaller than the maximum, whereas if we use the mean for the minimisation, we risk having high values of the PSB offset by low values.

The R function `twang` allows us to set other important parameters such as the maximum number of trees to be combined (to reduce the risk of over-fitting), their maximum interaction level, and the shrinkage level. In this work, we used mainly default values of the function `mnps` in the `R` package `twang`; except for the following cases, in which we also followed other suggestions found in literature(McCaffrey *et al.*, 2004) (the results of these additional attempts are available from the authors):

- The number of generalised boosted model iterations (`n.trees`): we used the default value (10,000) for the empirical application, but, since we observed that the balance was reached with fewer iterations, we set it at 3000 in order to save time and computational effort. However, we also ran also some simulations with 5000, 10,000, and 20,000 generalised boosted model iterations in order to check whether the lower number negatively affected the performance of inverse probability of treatment weighting.
- A shrinkage parameter was applied to each tree in the expansion (`shrinkage`). The default value was 0.01, but we also used 0.0005 in some simulations, as suggested in the literature(McCaffrey *et al.*, 2004).
- For the fraction of the training set, observations were randomly selected to propose the next tree in the expansion (`bag.fraction`), while introducing randomness into the model fit if it was less than 1. The default value was 1, but we also used 0.5 in some simulations, as suggested in the literature(McCaffrey *et al.*, 2004).
- For the maximum number of iterations for the direct optimisation (`iterlim`): the default value was 1000, but we also tried some simulations with a higher value (10000) to check whether 1000 was enough.

After the GBM computation of the propensity scores for each individual and for each treatment with respect to the rest of the population has been implemented, the result is a matrix with K propensity scores for each individual, each one of which is referred to as one of the treatments. In other words, the first part of the algorithm produces a matrix that shows the computed probability of living in each neighbourhood (and not in other neighbourhoods) for each subject. This step produces propensity scores that are useful for making each treatment group comparable with the rest of the population. The sum of the K propensity scores for each individual is not equal to 1, as in the multinomial model, because these values are the results of different models that consider treatments separately. Since propensity scores are used in this work primarily in order to balance the weights, this is not an issue, and it is not necessary to modify these values to make them sum to 1. Indeed, such a transformation would simply modify the scale of the weights, and would not have any effect on the final result. The final weight for each subject is given by the inverse of the propensity score of the received treatment, while all of the other weights computed for that individual and relative to the propensity for receiving other treatments are discarded.

When dealing with IPTW, it is common to find extremely high weights that cause the variance of estimates to increase. Therefore, weight trimming has been considered as a way to reduce the variance with small losses in terms of bias(Lee *et al.*, 2011). However, the optimal level of trimming for improving the inference and achieving the best compromise between bias and variance is difficult to determine. Thus, it is sometimes more effective to focus on the procedure for computing weights, such as a proper specification of the propensity score model (Lee *et al.*, 2011). This is the main reason why we also implemented an asymmetrical trimming of the higher weights in the simulation study, while setting the extreme weights equal to the upper bound threshold, even if there is no proof of a substantial improvement in the overall performance of the GBM in an inverse weighting procedure in dichotomous cases (Lee *et al.*, 2011).

4.3 Simulation design

In order to keep our experiment realistic and to simplify our computations, we extracted from the total population a 10% sample from each neighbourhood. The original data structure was thereby preserved, but with a reduced sample size that makes the computations less demanding (the simulation dataset contains 22,690 individuals). In order to keep the simulation simple, we selected a small number of covariates from the variables described in section 3.2.1: gender, age, education (Edu0, Edu1, Edu2, and Edu3), overcrowding, hypertension, and drugs.

In the simulation experiment, we included variables describing the health conditions of the population that we had discarded in the empirical study, because in our simulations we established both the temporality and the causality direction given by the data generation design: i.e., we simulated first the treatment and then the outcome; whereas in the empirical framework, this assumption could not be trusted with respect to health conditions.

In line with other studies(Arpino and Cannas, 2016; Setoguchi *et al.*, 2008), and given the real distribution of these six covariates, we decided to simulate the treatment assignment and the outcome according to three different scenarios that reflect three different treatment allocation settings: the first one reflects the real circumstances with a simple, linear, and additive model; the second one shows a case in which the treatment allocation equation is complex and may be misspecified; and the third one represents a highly unbalanced situation.

In the first scenario, the treatment assignment equation is simple and close to reality. The treatment is generated through a multinomial logistic model, using neighbourhood 6 as a reference; that is, the neighbourhood with the lowest crude hospitalised fractures rate. Thus, for each neighbourhood t and each individual i, the treatment equation will be

$$\begin{aligned} ln\left(\frac{Pr(T_i = t)}{Pr(T_i = 6)}\right) &= {}_1\beta_0^t +_1 \beta_1^t * Gender_i +_1 \beta_2^t * Age_i +_1 \beta_3^t * Edu1_i +_1 \beta_4^t * Edu2_i + \\ &+ {}_1\beta_5^t * Edu3_i +_1 \beta_6^t * Overcrowding_i +_1 \beta_7^t * Hypertension_i + \\ &+ {}_1\beta_8^t * Drugs_i. \end{aligned} \tag{4.5}$$

In order to choose the values for the coefficients, we estimated a multinomial logistic model on the whole population and used the same rounded parameters for $t = 1, ..., 5, 7, ..., 10$ for the intercept ${}_1\beta_0^t$ and for other coefficients ${}_1\beta_v$ $v = 1, ...8$ (the exact values of the parameters are reported in table A.1 in appendix A.1).

The second scenario relies on a more complex treatment assignment equation that includes six interaction terms and three quadratic terms, while having the following equation form for each neighbourhood t

$$\begin{aligned} ln\left(\frac{Pr(T_i = t)}{Pr(T_i = 6)}\right) &= {}_2\beta_0^t +_2 \beta_1^t * Gender_i +_2 \beta_2^t * Age_i +_2 \beta_3^t * Edu1_i +_2 \beta_4^t * Edu2_i + \\ &+ {}_2\beta_5^t * Edu3_i +_2 \beta_6^t * Overcrowding_i +_2 \beta_7^t * Hypertension_i + \\ &+ {}_2\beta_8^t * Drugs_i +_2 \beta_9^t * Age_i^2 +_2 \beta_{10}^t * Overcrowding_i^2 + \\ &+ {}_2\beta_{11}^t * Drugs_i^2 +_2 \beta_{12}^t * Gender_i * Age_i +_2 \beta_{13}^t * Gender_i * Cardio_i + \\ &+ {}_2\beta_{14}^t * Gender_i * Drugs_i +_2 \beta_{15}^t * Age_i * Cardio_i + \\ &+ {}_2\beta_{16}^t * Age_i * Drugs_i +_2 \beta_{17}^t * Drugs_i * Cardio_i. \end{aligned} \tag{4.6}$$

As in the first scenario, the parameters for these treatment assignment equations were chosen based on the parameters estimated by a multinomial logistic model with the same functional form for the whole population (the exact values of the parameters are reported in table A.2 in appendix A.1).

The third scenario relies on the very same treatment assignment equation as in the first scenario, but with different parameters. Indeed, starting with the coefficients in scenario 1, some of the parameters were modified to obtain a greater initial imbalance. Moreover, in order to keep the simulated dataset close to a potentially real situation in terms of the hospitalised fractures percentage, the intercepts were modified as well (the exact values of the parameters are reported in table A.3 in appendix A.1).

We evaluated the initial balance of these three scenarios in all of the 1000 simulations using the Population Standardized Bias. The mean values of the PSB among all of the 1000 simulations for each scenario are reported in table 4.1. While in the first scenario the initial situation is only mildly unbalanced, in scenarios 2 and 3 more extreme imbalanced situations can be observed.

Table 4.1: Mean of PSB among neighbourhoods of the unweighted sample in all the iterations

		Neighbourhoods									
Scenario	**Variable**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	Male	4.64	1.33	1.59	2.58	2.33	2.49	1.68	2.33	1.61	5.02
	Female	4.64	1.33	1.59	2.58	2.33	2.49	1.68	2.33	1.61	5.02
	Age	17.84	5.61	5.11	6.88	8.46	9.47	3.80	12.70	2.61	16.65
	Primary Educ. or lower	36.04	10.56	6.76	6.53	22.68	23.36	2.24	26.42	1.35	26.14
	Lower Secondary Educ.	8.40	5.94	3.04	1.74	2.68	3.29	1.30	3.76	5.42	4.94
	Upper Secondary Educ.	19.90	6.99	4.81	5.79	16.12	17.14	1.86	15.57	2.40	18.54
	Tertiary Educ.	31.30	1.56	0.83	1.14	15.87	14.45	1.27	20.77	5.61	15.36
	No Hypertension	7.41	1.48	1.83	1.77	2.22	5.29	1.91	4.23	1.80	3.95
	Hypertension	7.41	1.48	1.83	1.77	2.22	5.29	1.91	4.23	1.80	3.95
	Overcrowding	26.34	6.00	2.19	1.36	12.02	10.18	2.54	21.93	4.44	2.32
	Drugs	24.53	1.71	5.29	4.04	10.31	12.92	2.14	13.22	3.99	14.71
2	Male	6.26	1.41	1.64	6.05	1.42	6.38	6.22	3.76	2.78	1.59
	Female	6.26	1.41	1.64	6.05	1.42	6.38	6.22	3.76	2.78	1.59
	Age	30.81	6.14	13.49	29.99	2.49	14.87	26.13	26.09	38.62	1.58
	Primary Educ. or lower	29.91	9.77	10.60	8.64	23.39	14.35	2.73	21.47	4.02	20.31
	Lower Secondary Educ.	3.42	5.39	4.00	2.78	3.14	1.89	2.67	3.22	5.84	7.39
	Upper Secondary Educ.	15.61	6.47	6.36	7.04	15.14	10.56	2.21	13.87	5.84	12.81
	Tertiary Educ.	24.22	1.21	2.25	2.45	18.31	14.96	2.33	15.25	7.51	6.26
	No Hypertension	1.39	3.94	5.45	6.98	1.79	3.30	5.22	9.65	7.74	2.51
	Hypertension	1.39	3.94	5.45	6.98	1.79	3.30	5.22	9.65	7.74	2.51
	Overcrowding	19.63	9.68	2.65	10.62	17.49	11.54	11.88	14.70	15.59	1.00
	Drugs	13.52	11.52	9.99	11.93	3.85	2.52	15.86	17.31	7.14	4.79
3	Male	22.94	3.75	6.95	38.93	8.38	5.34	1.55	13.88	7.79	5.61
	Female	22.94	3.75	6.95	38.93	8.38	5.34	1.55	13.88	7.79	5.61
	Age	258.46	23.91	17.88	7.20	27.36	23.28	15.68	20.03	18.78	13.86
	Primary Educ. or lower	4.19	22.26	23.94	9.30	27.75	26.14	10.24	52.42	3.44	31.52
	Lower Secondary Educ.	6.70	33.45	10.18	4.36	2.88	2.24	6.12	27.87	8.68	5.90
	Upper Secondary Educ.	1.12	4.59	44.43	1.15	21.08	20.27	15.01	10.57	5.85	22.65
	Tertiary Educ.	2.29	12.91	13.03	9.53	21.48	19.93	8.85	92.91	13.00	21.24
	No Hypertension	28.29	3.94	3.70	2.42	1.16	3.22	7.56	10.45	1.59	13.86
	Hypertension	28.29	3.94	3.70	2.42	1.16	3.22	7.56	10.45	1.59	13.86
	Overcrowding	34.09	6.35	5.08	3.67	30.76	8.17	3.47	21.79	4.21	4.67
	Drugs	27.08	4.81	11.89	2.72	5.21	9.27	8.47	28.43	2.24	44.81

After the treatment generation, the outcome has also been simulated given the six covariates and the treatment assignment according to the following model:

$$\begin{aligned} ln\left(\frac{Pr(Y_i=1)}{Pr(Y_i=0)}\right) &= \beta_0+\beta_1*Gender_i+\beta_2*Edu1_i+\beta_3*Edu2_i+\beta_4*Edu3_i+ \\ &+ \beta_5*Cardio_i+\beta_6*Age_i+\beta_7*Overcrowding_i+\beta_8*Drugs_i+ \\ &+ \beta_9*D_{i1}(T_i)+\beta_{10}*D_{i2}(T_i)+\beta_{11}*D_{i3}(T_i)+\beta_{12}*D_{i4}(T_i)+ \\ &+ \beta_{13}*D_{i5}(T_i)+\beta_{14}*D_{i7}(T_i)+\beta_{15}*D_{i8}(T_i)+\beta_{16}*D_{i9}(T_i)+ \\ &+ \beta_{17}*D_{i10}(T_i), \end{aligned} \tag{4.7}$$

where $D_{i1}(T_i)$, $D_{i2}(T_i), ..., D_{i10}(T_i)$ are dichotomous variables that take value 1 if the individual i lives in the considered neighbourhood, and value 0 otherwise. As before, the reference is neighbourhood 6. The coefficients are close to those estimated by the same model for the whole population, but the parameters from β_9 to β_{17} were inflated to obtain a larger neighbourhood effect for the purposes of estimation (the exact values of the parameters are reported in table A.4 in appendix A.2). Indeed, when the true neighbourhood effects are small, there is a risk that the simulations will produce more biased and less stable estimates, and that the inverse weighting approach will perform badly (Cepeda *et al.*, 2003). . However, we also ran some simulations with smaller neighbourhood effects in order to explore and verify this result in a multi-treatment framework.

We evaluated the performance of the two approaches, the logistic regression, and the IPTW, while comparing the estimates of nine neighbourhood coefficients (the reference is neighbourhood 6, the one with the lowest crude rate of hospitalised fractures) and the true treatment effect used to simulate the outcome. The analysis was focused on three measures: the mean and the median of the relative bias (the percentage difference from the true treatment effect), the variance of the estimated values among the 1000 simulations, and 95% confidence interval coverage (the percentage of times the true value is included in the 95% confidence interval of the obtained estimates among all of the simulations). In this setting, we preferred to take into account the median, and not just the mean, of the bias because the median is less influenced by extreme values that may be the consequence of less plausible scenarios.

4.4 Simulation results

For each replicate in every scenario, we estimated the neighbourhood effect using both the logistic regression approach and the IPTW approach. Since we were trying to improve the balance of the confounders among the neighbourhoods, we observed the distribution of the PSB across all of the simulations, neighbourhoods, and variables. To summarise them all, we reported the mean of the PSB of the weighted samples among all of the simulations in table 4.2.

In the first two scenarios, the balance attained with IPTW was extremely good, with all of the considered confounders showing an average PSB that was lower than 5% for all of the neighbourhoods; indeed, in many cases, the PSBs was even lower than 1%. Using even the most restrictive threshold cited in literature, we can state that in these two scenarios the balance was reached.

In the most complicated scenario (scenario 3), the PSBs tended to be higher. This was especially the case for neighbourhood 1, for which most of covariates had PSBs higher than 10%, and the average PSB for age was 53.23%. Even though the balance was not satisfactory, it should be noted that in scenario 3 the initial imbalance was very high (e.g., the PSB for age in neighbourhood 1 was 258.46; table 4.1). Indeed, if we compare the balance after weighting (table 4.2) with the initial balance, we can see that even in scenario 3, the use of the IPTW approach guarantees a considerable improvement in the degree of similarity of the confounders' distributions across the neighbourhoods. Since the residual imbalance was higher, we expected to observe higher bias for the IPTW estimator in scenario 3.

Whereas in scenarios 1 and 2 the bias of the IPTW estimates was quite good, or lower than 5% in most cases; in the third scenario, there were two parameters with a bias higher than 10%. However, as was already explained, in the third scenario, the initial balance was particularly challenging in terms of the distribution of the confounders among the different treatment groups. Moreover, when the bias of IPTW estimates was high, the logistic regression method also provided biased estimates.

In the first scenario, we observe that the biases relative to the estimates produced by IPTW were smaller than those produced by the logistic regression method, except for one neighbourhood, number 10. This neighbourhood had the highest PSBs, and was the only one for which a PSB higher than 5% was found.Relative to the other neighbourhoods, the third had the largest bias with respect to both the estimation approaches and the mean and the median. Indeed, the bias of this parameter was expected to be the highest because its true value was the smallest and closest to 0. According to the literature(Cepeda *et al.*, 2003), higher bias is often observed for estimates of smaller effects. In general, in the first scenario, almost all of the parameters were estimated by

TABLE 4.2: Mean of PSB among neighbourhoods of the weighted sample in all the replicates

		Neighbourhoods									
Scenario	**Variable**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	Male	1.90	0.42	0.51	0.73	0.82	0.88	0.54	1.42	0.48	1.62
	Female	1.90	0.42	0.51	0.73	0.82	0.88	0.54	1.42	0.48	1.62
	Age	0.96	1.45	0.79	1.02	2.66	2.91	1.30	1.47	2.00	5.48
	Primary Educ. or lower	1.55	0.29	0.27	0.40	1.60	1.67	0.56	1.01	0.57	3.23
	Lower Secondary Educ.	0.83	0.40	0.32	0.40	0.53	0.51	0.38	0.84	0.72	0.80
	Upper Secondary Educ.	0.73	0.24	0.20	0.26	0.72	0.87	0.50	0.41	0.42	1.75
	Tertiary Educ.	0.35	0.53	0.27	0.33	2.38	2.26	0.55	0.27	1.45	4.04
	No Hypertension	0.93	0.42	0.31	0.44	0.65	1.01	0.74	0.91	0.64	1.53
	Hypertension	0.93	0.42	0.31	0.44	0.65	1.01	0.74	0.91	0.64	1.53
	Overcrowding	3.35	1.05	0.26	0.41	0.47	0.41	0.43	3.12	0.45	2.10
	Drugs	1.08	0.96	0.52	0.80	2.18	2.41	0.98	1.24	1.69	4.48
2	Male	1.04	0.42	0.85	2.67	0.67	0.70	1.20	2.35	1.73	0.60
	Female	1.04	0.42	0.85	2.67	0.67	0.70	1.20	2.35	1.73	0.60
	Age	0.52	0.53	1.87	6.13	1.73	2.88	4.59	6.21	7.87	2.01
	Primary Educ. or lower	1.07	0.35	0.62	1.23	1.64	1.62	1.24	1.56	2.35	1.22
	Lower Secondary Educ.	0.66	0.31	0.51	1.13	0.63	0.63	1.09	1.26	1.80	0.19
	Upper Secondary Educ.	0.38	0.22	0.38	0.91	0.59	0.71	1.29	0.86	2.00	1.02
	Tertiary Educ.	0.22	0.21	0.31	1.09	2.78	2.58	1.00	0.68	3.62	0.63
	No Hypertension	0.57	0.35	0.59	1.42	0.43	0.96	1.27	1.56	1.52	0.50
	Hypertension	0.57	0.35	0.59	1.42	0.43	0.96	1.27	1.56	1.52	0.50
	Overcrowding	1.33	1.93	0.50	0.94	0.48	0.40	0.79	3.22	1.23	0.48
	Drugs	0.57	1.18	0.63	2.23	1.64	1.97	1.64	2.28	3.55	0.97
3	Male	9.59	0.85	0.57	1.66	1.53	1.46	0.98	1.47	1.21	2.25
	Female	9.59	0.85	0.57	1.66	1.53	1.46	0.98	1.47	1.21	2.25
	Age	53.23	5.51	4.00	4.59	7.35	7.11	5.47	4.82	5.87	8.70
	Primary Educ. or lower	10.46	0.54	0.33	1.28	2.80	2.59	1.73	1.73	1.36	5.18
	Lower Secondary Educ.	8.80	1.30	0.58	0.61	0.73	0.81	0.69	0.88	1.13	1.13
	Upper Secondary Educ.	9.16	0.45	0.54	0.74	1.20	1.15	1.79	0.81	0.85	2.54
	Tertiary Educ.	7.07	2.19	1.80	2.32	4.02	3.86	1.45	0.70	2.76	6.74
	No Hypertension	11.19	0.61	0.54	0.78	0.78	0.97	0.99	1.54	0.73	2.61
	Hypertension	11.19	0.61	0.54	0.78	0.78	0.97	0.99	1.54	0.73	2.61
	Overcrowding	18.68	1.36	0.49	0.85	1.98	0.66	0.64	2.02	0.59	2.20
	Drugs	17.18	1.22	0.74	1.07	2.31	2.62	1.24	1.91	1.87	9.59

both of the models with bias lower than 5%; except for neighbourhoods 1, 3 (already mentioned), and 4, where, on average, the inverse weighting approach seems to have provided better estimates.

In the second scenario, the estimates given by the IPTW method for neighbourhoods 4, 8, and 9 had a particularly high median bias, of between 5% and 10%. This was probably because the balance in these neighbourhoods was not completely achieved, especially for the variable age, which had a mean PSB of more than 5% in these three neighbourhoods (table 4.2). On the other hand, the logistic regression model provided estimates that were particularly biased for the effect of neighbourhood 1; probably because of the initial highly unbalanced situation (as shown in table 4.1).

In the third scenario, both methods performed well for most of the neighbourhoods (numbers 2, 5, 7, 8, 9, and 10), as they had both mean and median biases of less

TABLE 4.3: Variance, 95% Confidence Interval Coverage and bias (mean and median) of parameters' estimates in the 3 scenarios, comparison between regression and inverse probability of treatment weighting adjustment (IPTW)

		Scenario 1				Scenario 2				Scenario 3			
Neigh.		Bias		Var	95% CI	Bias		Var	95% CI	Bias		Var	95% CI
(*true value*)	Method	Mean	Median		Coverage	Mean	Median		Coverage	Mean	Median		Coverage
1	Logit	6.09	5.26	0.08	96.0	13.32	10.64	0.08	94.2	62.38	60.64	0.09	63.2
(0.820)	IPTW	3.90	3.74	0.10	96.1	6.28	3.93	0.09	95.4	37.16	27.75	0.29	88.3
2	Logit	2.05	1.29	0.07	95.3	2.85	1.20	0.07	96.2	3.09	2.13	0.10	95.5
(1.310)	IPTW	0.97	−0.15	0.08	95.0	1.67	0.33	0.07	96.4	2.66	0.77	0.15	93.0
3	Logit	11.92	11.96	0.08	95.8	3.28	3.67	0.11	95.0	12.97	9.80	0.11	95.9
(0.375)	IPTW	11.19	8.24	0.09	94.9	4.96	3.84	0.12	94.9	18.30	14.17	0.17	93.2
4	Logit	5.79	3.07	0.08	96.2	−4.87	0.32	0.22	96.6	3.75	2.76	0.11	95.2
(0.720)	IPTW	4.40	2.39	0.10	95.4	−9.53	−5.37	0.25	96.3	8.22	5.57	0.16	92.7
5	Logit	2.67	2.36	0.08	95.8	3.86	2.71	0.08	95.7	2.03	1.38	0.11	94.5
(0.915)	IPTW	1.59	1.41	0.09	95.0	1.43	0.08	0.09	94.1	1.16	−0.78	0.17	92.1
7	Logit	2.37	1.27	0.07	95.7	−0.71	−0.34	0.13	95.3	2.59	1.59	0.10	95.9
(1.430)	IPTW	0.71	−0.02	0.08	95.0	−4.25	−4.01	0.14	95.0	1.98	−0.02	0.15	92.5
8	Logit	3.66	4.33	0.09	96.5	−3.12	−0.28	0.21	94.5	5.18	3.81	0.12	95.6
(0.950)	IPTW	−0.12	−1.21	0.11	95.8	−7.18	−4.63	0.24	93.7	−0.29	−0.15	0.22	94.2
9	Logit	3.01	3.14	0.08	96.1	−5.78	−2.43	0.22	96.2	1.51	−0.58	0.11	96.4
(1.020)	IPTW	1.36	−0.32	0.09	95.7	−11.52	−8.50	0.26	95.1	1.82	−0.51	0.15	94.0
10	Logit	−0.54	−1.13	0.09	95.8	1.36	0.06	0.07	94.8	1.17	−0.51	0.11	94.8
(1.535)	IPTW	−3.68	−3.96	0.10	95.1	−0.59	−1.20	0.08	94.1	−0.53	−1.81	0.18	93.0

than 5%. However, when the IPTW approach was applied, the biases became slightly smaller for all of these neighbour- hoods. The logistic regression model produced better results in terms of bias in neighbourhoods 3 and 4. However, both methods performed poorly with respect to the estimates for neighbourhood 1, for which the initial situation was extremely unbalanced (as shown in table 4.1). While the situation of the first neighbourhood remained unbalanced even after weighting (as shown in table 4.2), the IPTW approach produced estimates for this parameter that had, on average, half the bias of those produced by the logistic regression model.

A general observation with respect to table 4.3 is about variances. Indeed, in all of the scenarios and for all of the neighbourhoods, the variances of the estimates generated by the inverse probability of treatment weighting approach were higher than those produced by the logistic regression model. In the first scenario, which corresponds most closely to reality, the variances of the two models were more similar and smaller than those in the other two scenarios, in which the allocation of individuals to treatments was more complex (in the second one) and more unbalanced (in the third one).

Since in the presence of weights the variance may increase and the estimates may be greatly affected, especially if the weights are extreme, we also tried an asymmetrical trimming. We trimmed only the extremely high weights, reducing the influence of those individuals who were under-represented in some of the neighbourhoods, based on the assumption that these individuals were outliers who did not reflect the population as a whole. Selecting different levels of trimming (percentiles from 99 to 85 and 75), we assigned the threshold weight value to those individuals who had higher weights. This technique proved to be quite useful for reducing the variance, but the gain was associated with an increase in bias in some cases.

In figures 4.1 and 4.2, two examples are presented that show how trimming affected the mean bias and the variance of the estimates produced by the IPTW approach relative to the mean bias and the variance of the logistic regression's estimates for neighbourhoods 5 and 8 in the first scenario. In some cases, as for neighbourhood 5 in figure 4.1, it was possible to increase the level of trimming while having a limited impact on the bias, or even causing it to decrease slightly at around the 85^{th} percentile, while ensuring that it remained lower than the bias of the estimates produced by the logistic regression. On the other hand, the variance of the trimmed estimates was substantially reduced, and assumed values closer to the variance of the logistic regression's estimates when the level of trimming was increased. Thus, when we consider this example, we can state that the optimal level of trimming in order to reduce both the bias and the variance may be around the 85^{th} percentile. A completely different situation can be

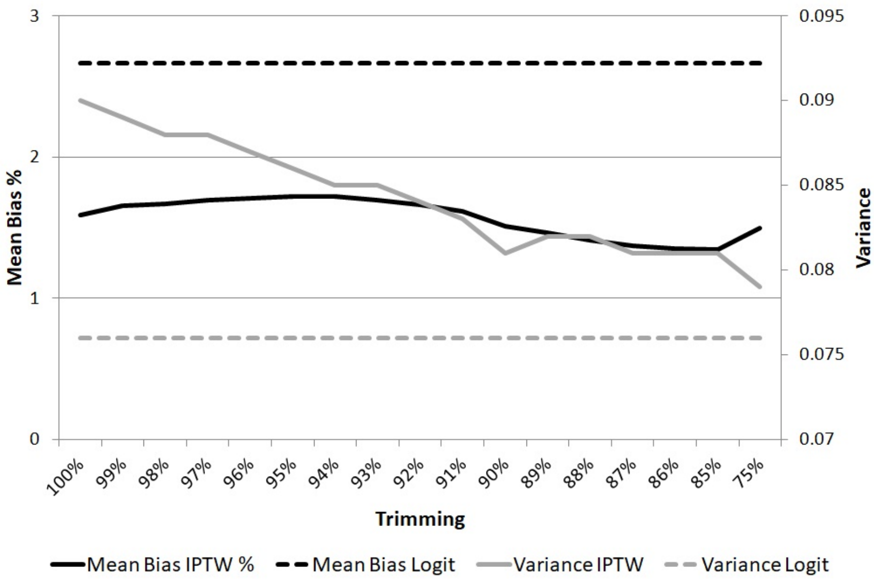

FIGURE 4.1: Comparison of mean biases and variances of estimates obtained by IPTW at different levels of trimming for the neighbourhood effect of neighbourhood 5 in the first scenario with logistic regression's (Logit) results (represented as horizontal dashed lines).

observed for the estimates of neighbourhood 8 in the first scenario, where the trade-off between the bias and the variance was more severe than in the previous example. Indeed, when the level of trimming was increased, the bias grew from around 0% in the absence of trimming to around 8% at the 75^{th} percentile of trimming. However, the variance decreased when moving closer to the variance of the logistic regression's estimates. Indeed, finding the optimal level of trimming was harder in this case, as the level at which the estimates were less biased was the one at which the variances were higher. Moreover, at around the 96^{th} percentile of trimming, we got estimates with the same bias as the logistic regression's estimates, but with a variance that was 15% higher.

In general, even after observing all of the trimmed estimates in all of the simulations, it was not possible to find a common criterion we could use to define a best practice in terms of trimming. The fact that we had nine different parameters to estimate did not make this choice easier, because the levels that ensure a balance between bias and variance may be different for each parameter. Moreover, in an empirical framework, it is not possible to observe the bias of estimates. Thus, it would have been even more difficult to discern which trimming level was the best to use without quantifying the loss in terms of bias. Thus, we would not recommend the trimming of weights when using the IPTW approach in a multi-treatment framework.

In order to improve the performance of the inverse probability of the treatment weighting approach, we also tried to change some default settings in the twang package.

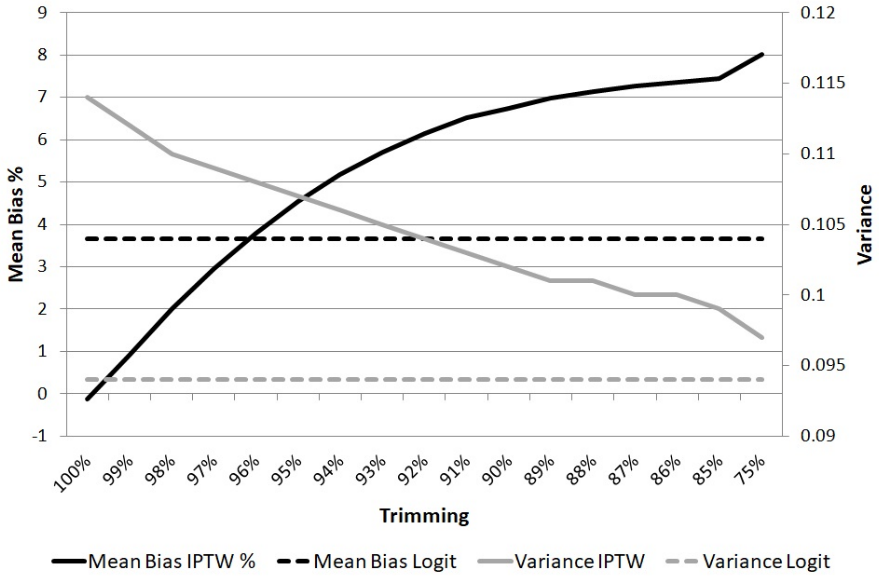

FIGURE 4.2: Comparison of mean biases and variances of estimates obtained by IPTW at different levels of trimming for the neighbourhood effect of neighbourhood 8 in the first scenario with logistic regression's (Logit) results (represented as horizontal dashed lines).

As we mentioned before, we ran some simulations with different numbers of generalised boosted model iterations (3000, 5000, 10,000, and 20,000), levels of shrinkage (0.01 and 0.0005), fractions of the training set to fit the trees (1 and 0.5), and maximum numbers of iterations for the direct optimisation (1000 and 10000); as well as several combinations thereof. However, as the balance after weighting was not improved and the bias was not reduced, we decided against deepening this research path, and instead opted to use all of the default values for the simulations, except for the number of generalised boosted model iterations (the default was 10,000, but to save time and computational effort, we used 3000, since the balance was reached with fewer iterations).

4.5 Conclusions

The purpose of this chapter was to estimate neighbourhood effects adjusting for confounders through IPTW techniques, using as a motivating study the estimation of neighbourhood effect on hospitalized fractures in the Italian city of Turin. One of the most intriguing points with respect to a methodological point of view is linked to the number of treatments that is not trivial to handle. The main objective of this chapter was to examine the performance of IPTW approach in the case of many treatments (10, specifically). This was done implementing simulation studies where IPTW was also compared to a standard logistic regression approach. These approaches were also applied on real data originating from our motivating case study.

The simulation study was performed under three possible scenarios for the allocation of individuals to neighbourhoods, one close to reality, one with a complex misspecified treatment allocation and one with an extremely unbalanced initial situation. IPTW performed very well in terms of reducing the initial imbalance of confounders across the different neighbourhoods in all scenarios. However, in scenarios characterized by higher initial imbalance, bias of estimated causal effect was higher than the first scenario characterized by a lower initial imbalance.

It is widely stressed in the causal inference literature (Cannas and Arpino, 2018) that researchers should examine balance measures because higher (residual) imbalance tend to be associated with higher bias of causal estimates. This is confirmed in our analyses where bias tended to be higher for both logistic regression and IPTW in scenarios characterized by higher initial (and residual) imbalances.

Our results indicate that IPTW is a promising approach for reducing confounders imbalance in a multi-treatment context even in the presence of a number of treatments as high as 10.

However, the IPTW approach is more computationally demanding than a standard logistic regression (computation of weights may last several hours if the number of treatments is high, as in our case). Future research may be devoted to investigate more computationally efficient approaches, which are especially necessary in the presence of a even higher number of treatments than we considered here.

One limitation in the application we considered is represented by the fact that mobility of individuals among neighborhoods may invalidate the SUTVA. Indeed, in some neighbourhoods individuals may have a higher propensity to move and to be affected by other neighbourhoods. Since our focus is on older people that is considered a more stable population, this risk is lower, but in future researches would be interesting to take into account also this aspect.

Chapter 5

Neighbourhood effect with an IPTW approach

5.1 Introduction

The increase in average life expectancy during the 20th century ranks as one of society's greatest achievements (WHO *et al.*, 2011). Indeed, data on life expectancies between 1840 and 2007 show a steady increase averaging about three months of life per year. In particular, according to Istat, the life expectancy at birth in Italy has grown from 67.2 years for men and 72.3 for women in 1961 to 80.1 years for men and 84.7 for women in 2015; according to Istat projections, these numbers are going to increase to 86.6 years for men and 91.5 for women in 2065 (Istat, 2016). However, health inequalities are still present in the territory. These are mostly due to socio-economical differences, but also to the geographical residence (Graham, 2009). During the last years, one important objective of health policies has been the reduction of health inequalities, with an increased attention to elderly people. With this aim, also the investigation of neighborhood effect on health outcomes earned attention in social epidemiology.

5.2 Neighbourhood effect on two health outcomes

In this chapter some empirical results are reported, the neighbourhood effect has been estimated with IPTW method described in chapter 4 on two different health outcomes: the hospitalized fractures and mental health. These two were selected because of their importance among old population but also because the neighbourhood may affect them in two different ways. The hospitalized fracture may be seen more as an event than as a disease, thus it seems plausible that previous events and history of individuals have a low impact on the probability of the outcome. Thus, the hypothesis that just the neighbourhood in which each subject is living needs to be taken into account and not the previous holds. On the contrary, mental diseases are for sure the results of a complex mixture of events and elements that intersect with the time dimension.

According to characteristics of the two outcomes, we selected the populations for the analysis in order to make the causal inference assumptions described in section 2.2 hold.

The geographical partition that we consider in this chapter counts just 10 neighbourhoods, indeed, since the IPTW technique proposed by McCaffrey *et al.* (2013) is computationally demanding, the computation of neighbourhood effect on neighbourhoods with smaller scale would be highly unpractical and nearly impossible.

5.3 Neighbourhood effects on hospitalized bones fractures

Interest in the neighbourhood effect on hospitalized fractures among over-60-year-olds stems from a real need expressed by Turin's Epidemiology Service. Neighbourhoods may affect elderly fracture rates in two main ways: they may be difficult to walk around, or have inadequate street lighting, and thus increase the risk of falls; and/or people living in the area may be discouraged from engaging in physical activity, and their muscle tone and bone structure consequently deteriorate (Ambrose *et al.*, 2013; Barnett *et al.*, 2017; Sànchez-Riera *et al.*, 2010). The focus here is on people over sixty, partly because of their greater exposure of hospitalized fracture, and also because they are assumed to be a more stable resident population. In fact, some researchers have found older people more susceptible to neighbourhood effects because they spend more time in their neighbourhoods than younger people (Melis *et al.*, 2015; Turrell *et al.*, 2014). Older people are also less likely to move house (the annual rate for the observed population was only around 1%).

We estimated neighbourhood effect using the IPTW method proposed by McCaffrey *et al.* (2013) and described in chapter 4. As in chapter 4, we computed the neighbourhood effect also with the logistic regression in order to compare the results.

5.3.1 Data and population

The analysed population consists of all participants in the 2001 population census, with some additional restrictions. We consider only the individuals who where aged 60 or older on 31 December 2001. In order to be able to collect information on possible confounders related to past health information, we focus on the individuals who were living in Turin between 1 January 1997 and 31 December 2001. Finally, we measure the outcome; i.e., the incidence of hospitalised fractures during the year following the census (2002). Therefore, we restrict our analyses to individuals who were living in Turin over the whole period between 1 January 1997 and 31 December 2002. Our design allows us to measure the time-varying confounders before the treatment, which is in turn measured before the outcome is observed. We excluded also individuals living in institutions such as rest or nursing home.

In this empirical application we consider the geographical partition given by the 10 districts.The list of neighbourhoods with their population (% of the total population of Turin in parentheses) is as follows (hereafter, the neighbourhoods are referred to using the corresponding numerical identifier):

1. *Centro, Crocetta*: 18,224 individuals (8.07%);
2. *Santa Rita, Mirafiori nord*: 30,437 individuals (13.48%);
3. *San Paolo, Cenisia, Cit. Turin, Pozzo Strada*: 33,072 individuals (14.64%);
4. *San Donato, Parella, Campidoglio*: 23,065 individuals (10.21%);
5. *Borgo Vittoria, Madonna di Campagna, Lanzo, Lucento, Vallette*: 30380 individuals (13.45%);
6. *Regio Parco, Barca, Bertolla, Barriera di Milano, Rebaudengo, Falchera, Villaretto*: 25,288 individuals (11.20%);
7. *Aurora, Vanchiglia, Sassi, Madonna del Pilone*: 20,434 individuals (9.05%);
8. *Borgo Po, San Salvario, Cavoretto*: 13,591 individuals (6.02%);
9. *Nizza Millefonti, Lingotto, Filadelfia*: 20,729 individuals (9.18%);
10. *Mirafiori sud*: 10,608 individuals (4.70%).

Thus, the aim of this section is to estimate the causal effect of living in a given neighbourhood at the 2001 census on the probability of suffering from at least 1 hospitalized fracture in 2002. The considered confounders are age, gender, region of birth, family composition, education level, last observed professional condition, home ownership and overcrowding, described in detail in section 3.2.1. Table 5.1 presents some descriptive statistics on the outcome and confounders considered in our empirical analyses by neighborhoods. In the first row of table 5.1 is reported the percentage of hospitalized fractures in each neighbourhood. This event affects a small portion of population, 0.9% of over-60 years old individuals living in Turin, with some differences among neighbourhoods: from neighbourhood 6 that presents the lowest proportion (0.71%) to the neighbourhood 1 with 1.05% of hospitalized fractures in 2002.

TABLE 5.1: Descriptive statistics on the outcome and confounders by neighbourhoods.

	Neighbourhoods										
Variables	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Total**
Hospitalized Fractures (%)	1.05	0.93	0.85	0.92	0.84	0.71	1.03	1.02	0.92	0.85	0.90
Female (%)	60.60	57.40	58.88	59.50	56.75	56.67	58.67	58.97	57.48	55.07	58.63
Age (***Mean***)	71.99	70.63	71.22	71.35	70.48	70.43	71.14	71.68	70.87	70.02	70.96
Region of Birth(%)											
Piedmont	56.43	48.84	50.12	49.59	34.74	34.92	48.73	59.47	47.64	30.75	45.93
North of Italy	13.83	14.75	15.14	15.63	13.56	13.09	12.67	13.43	14.80	12.87	14.12
Center of Italy	3.74	3.51	2.76	2.89	2.59	2.73	2.57	3.24	3.24	2.42	2.97
South of Italy	21.19	27.54	26.73	27.04	39.98	41.91	31.25	19.59	27.42	47.62	30.93
Outside of Italy	4.81	5.37	5.25	4.86	9.13	7.36	4.77	4.28	6.91	6.33	6.05
Family composition (number of components) (%)											
Alone (1)	35.74	26.46	30.05	31.09	25.89	26.44	31.34	32.20	27.37	20.65	28.73
Married couple (2)	33.99	44.97	42.33	41.30	44.90	43.62	41.00	37.97	43.46	45.71	42.28
Married couple (> 3)	17.35	19.14	17.24	16.85	18.55	19.52	17.19	18.20	19.15	23.34	18.41
No married couple (> 2)	12.92	9.42	10.38	10.76	10.66	10.42	10.47	11.63	10.02	10.29	10.58
Educational attainment (%)											
Primary or lower	26.05	40.73	43.04	43.15	60.99	61.42	48.37	31.42	47.19	63.03	46.94
Lower Secondary	25.73	34.15	32.40	31.43	29.22	28.84	30.53	28.47	33.88	27.95	30.69
Upper Secondary	25.43	18.38	17.28	17.80	7.64	7.23	14.03	22.96	13.70	6.73	14.94
Tertiary	22.79	6.74	7.28	7.61	2.15	2.51	7.08	17.16	5.23	2.29	7.42
Home owner (%)	71.15	81.43	77.99	75.48	71.21	72.54	76.87	78.99	80.01	79.48	76.34
Last observed professional condition (%)											
No observed work	13.75	11.61	13.28	14.11	14.25	15.72	16.65	13.92	14.74	12.08	14.00
Home-maker	34.05	35.24	36.02	34.81	35.02	33.74	33.51	34.85	33.50	36.53	34.74
Entrepreneur	16.90	5.75	6.73	7.07	2.45	2.34	6.09	13.63	4.89	1.73	6.34
White collars	24.45	26.73	24.53	24.67	17.32	17.14	22.79	24.75	23.29	14.88	22.33
Manual workers	10.85	20.66	19.44	19.33	30.96	31.06	20.96	12.85	23.59	34.78	22.59
Overcrowding (***Mean***)	0.64	0.74	0.78	0.77	0.84	0.82	0.78	0.66	0.79	0.76	0.77
Hypertension	51.16	54.89	54.49	54.73	57.04	58.85	56.52	53.09	56.45	58.08	55.58
Drugs (***Mean***)	7.35	7.84	7.72	7.78	8.09	8.16	7.87	7.61	7.96	8.18	7.86

5.3.2 Results of IPTW approach and comparison with logistic regression estimates

We estimated the weights for the IPTW approach using the default values of the function `mnps` in the `R` package `twang`, and included the `n.trees` number of generalised boosted model iterations that was set to 10,000, even if such a large number was not necessary. Indeed, as 5.1, shows, gains in balance become smaller after 3000 iterations; and in some cases, such as in neighbourhood 7, increasing the complexity of the GBM model may worsen the obtained balance of the weighted variables and cause overfitting. However, when a huge number of treatments are to be considered, the decision about what number of iterations are optimal for getting good results in correspondence with each of them it is not trivial.

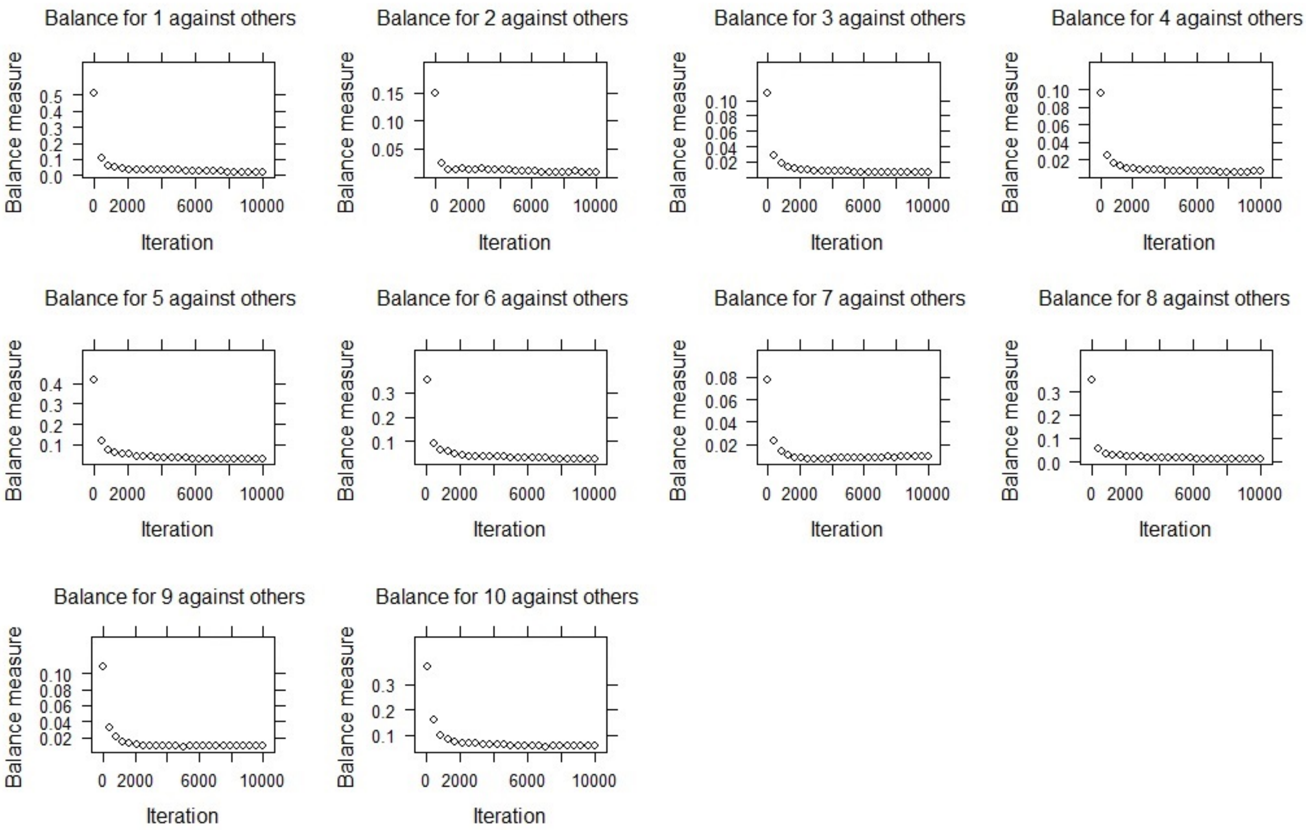

FIGURE 5.1: Reduction of the balance measure (the Population Standardized Balance, computed as in equation 4.4) during the weights estimation process in correspondence of an increasing number of iterations for the ten considered neighbourhoods, considering the whole population.

Nevertheless, as shown in figure 5.2, the final result is quite satisfying with respect to the initial unweighed situation, with almost all significant reduction of PSB differences considering the maximum among all pairwise comparisons.

There are small differences between the logistic regression and the IPTW estimates. We computed the neighbourhood effect for the whole population (All), and for the female (Women) and male (Men) populations separately. In table 5.2, we report the neighbourhood effect estimates in terms of the odds ratio, with standard errors and 95%

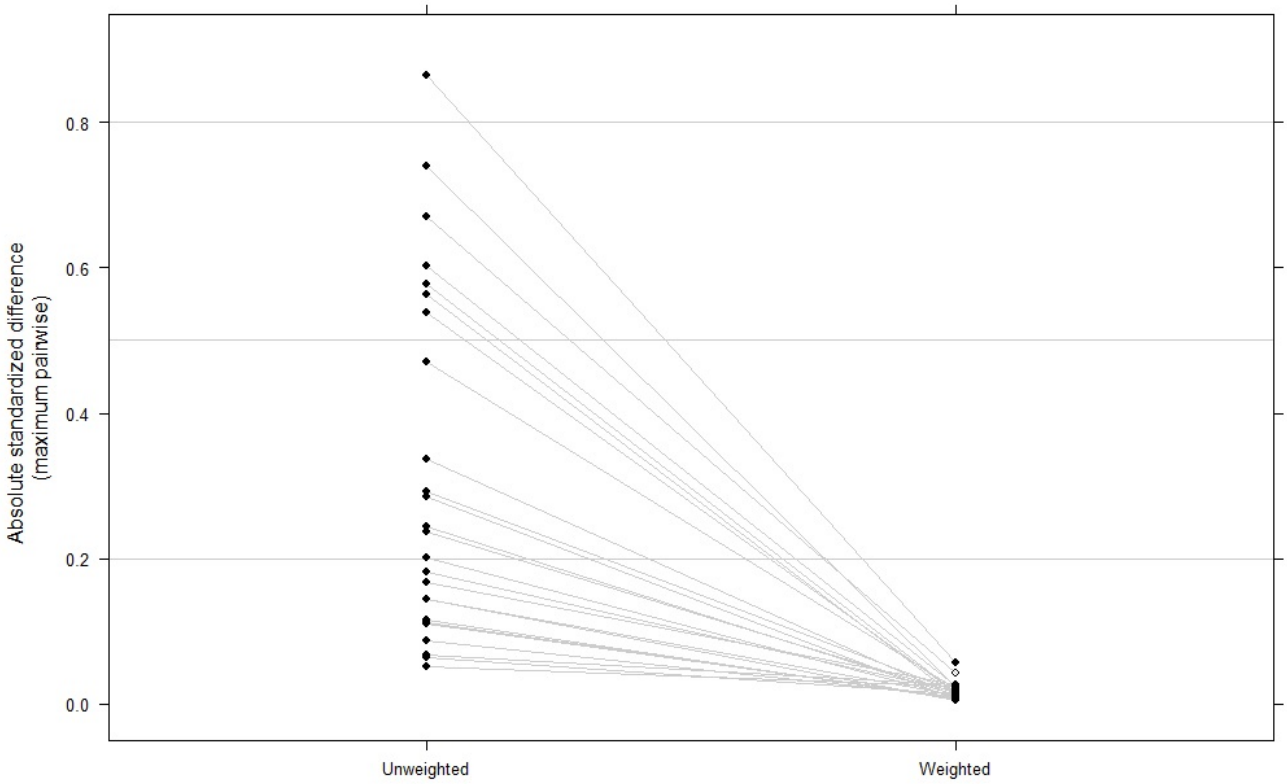

FIGURE 5.2: Comparisons of absolute standardised differences (considering the maximum of pairwise comparisons) in the whole population before and after weighting.

confidence intervals for the two estimation approaches, and with respect to the three different populations selected.

The parameters estimated with the two approaches on the whole population were similar, except for neighbourhood 10, for which the effect was greater in the logistic regression model. The odds ratio for the individuals living in neighbourhoods 2 and 7 was 33% higher than for the people living in neighbourhood 6 (odds ratio equal to 1.33). The main differences in the estimation of the neighbourhood effects were also observed in neighbourhoods 1 and 10 for the female population and in neighbourhoods 4, 5, 7, and 10 for the male population. In general, there were more discrepancies between the two estimation approaches for the effect of neighbourhood 10 than for the effects of the other neighbourhoods.

This may be explained with the observation of table 5.1, indeed the composition of this neighbourhood is quite different from the others. In this neighbourhood there is a higher percentage of men, individuals born in the South of Italy, subjects with primary or lower education than in the rest of Turin. Moreover, the most common last occupations are home-makers and labourers. Neighbourhood 10 is called "Mirafiori

Sud" and it is a very famous industrial area that started being densely populated in the 1950s when the FIAT factory was enlarged beckoning manpower from the South of Italy. The negative effect of this neighbourhood on health conditions of individuals living there is overestimated by the logistic regression probably because of a selection bias issue. Indeed, the difference between the two models' estimates is even greater for the male population.

Finally, it is interesting to notice that neighbourhoods have different impact on hospitalized fractures according to gender of individuals. While neighbourhood 6 is the area with lowest risk to experience the outcome if we consider both the entire and the female population, having in mind the male populations, neighbourhoods 10, 1 and 3 have a protective effect with respect to neighbourhood 6.

It is not possible to know exactly which of the two methods was more accurate in this setting, but, given the results of the simulation study, we can assume that the estimates based on the IPTW were more reliable because in the scenario closest to the real situation, this method performed better, with less bias.

5.4 Neighbourhood effects on mental health

Another interest of the SCaDU Services regards the neighbourhood effect on mental health, indeed it is involved also in an European project called *MINDMAP, Promoting Mental Wellbeing In The Ageing Urban Population: Determinants, Policies and Interventions In European Cities.* According to SCaDU Services, we decided to represent mental health on old individuals with two diseases, depression and dementia, that are the most common in over-60 years old population. Moreover, with the progressive aging of the population, as mentioned in the introduction, prevention and management of these degenerative diseases represent a strategical point in health policies planning.

In order to identify individuals affected by these diseases, we linked informations coming from different sources and detected idividuals affected by depression and dementia according to the following criteria:

Depression: at least on hospitalization with depression (source: hospital discharge records) or at least three drugs prescriptions of antidepressants (source: territorial drug prescriptions);

Dementia: at least on hospitalization with dementia (source: hospital discharge records) or at least three drugs prescriptions of anti-dementia drugs (source: territorial drug prescriptions) or having the exemption for dementia (participation in prescription charges).

Mental diseases are chronic and degenerative diseases, so it is difficult to identify their starting point and their determinants that caused the disease. Thus, we decided to analyse the incidence of these two diseases between 2002 and 2006, in order to be able to identify causes and estimate the neighbourhood causal effect.

5.4.1 Data and population

The analysed population consists of all participants to the 2001 population census with some additional restrictions. We consider only individuals aged 60 or more on 31^{st} December 2001. Moreover, we considered only individuals that had permanent residence in the same section in Turin from 1991 to 2001; this restriction guarantees that the estimated neighbourhood effect is referred to the observed neighbourhood in 2001 and not to other neighbourhoods. Moreover, we compared the distribution of confounders among discarded individuals and those included in the analysed population in order to verify that this restriction does not introduce a new source of selection bias. Since the outcome of interest is the incidence of mental diseases, all individuals that were already affected by dementia or depression before the 2002 (in the observation period between 1997 to 2001) are removed from the analysed population. We excluded also individuals living in institutions such as rest or nursing home. In the end, the population counts 194400 subjects.

As confounders we selected gender, age, ethnicity, educational attainment, marital status, last observed professional condition, living alone, home owner, type of housing and overcrowding. Moreover, in literature, some other diseases are listed as risk factors for the incidence of mental diseases such as hypertension, heart diseases and ictus, diabetes, Parkinson disease, disability and difficulty in movement, hypothyroidism and malignant tumor. Since these diseases may also be affected by the neighbourhood, it is not possible to use them in the propensity score computation.

5.4.2 Results and evaluation of the chronic disease variables' impact

In the first column of table 5.3 we reported the standardized rates of dementia and/or depression for sex and age in each of the 10 neighbourhoods. The neighbourhood with the lowest standardized rate is district 5, while the one with the highest standardize rate is district 10.

We computed the neighbourhood effect with the IPTW approach described in chapter 4 for the whole, the female and the male popoulation (reported in table 5.3 in columns called (1)). The estimated neighbourhood effect for the whole population, represented in figure 5.3, is expressed in terms of odds ratio and is significantly different from 1 at 5% level in three neighbourhoods: district 5 is the only one that has a protective effect (consistently with the standardized rate), while districts 3 and 4 have an odds ration higher than 1.

In order to include chronic diseases in the analysis we tried to include the six variables that represent risk factors for the outcome in the weighted logistic regression model together with the neighbourhoods' dummy variables. Our purpose is to verify if the estimated neighbourhood effect with IPTW is picking also up some other effects regarding health conditions of individuals that we did not include in the propensity score. On the contrary, effects, after the inclusion of chronic diseases variables, are more emphasized than in the model with just neighbourhoods' dummy variables.

Neighbourhood effects that are significantly different from 1 are not the same for the male and the female population, as already noticed in the analysis in section 5.2. Moreover, also the estimated effect with respect to the hospitalized fractures and the mental diseases are not exactly the same. This fact agrees with Galster (2008) theory that neighbourhoods may affect individuals' outcomes through different mechanisms that may interact differently with different health outcomes.

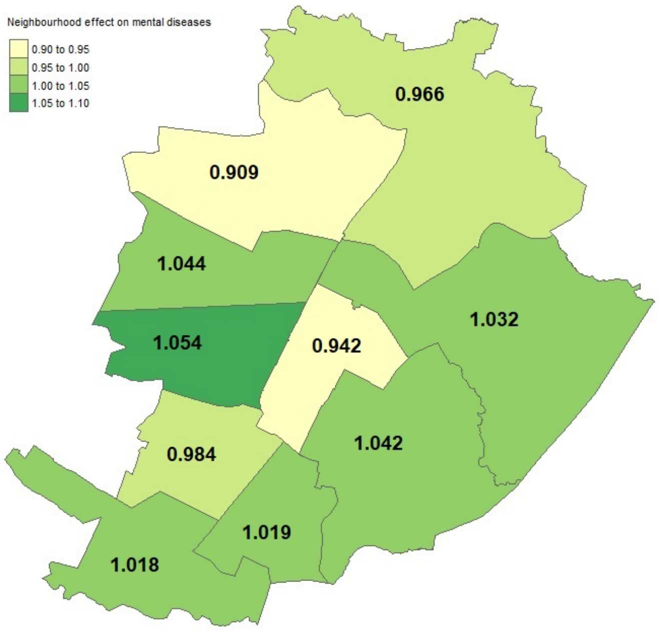

Figure 5.3: Neighbourhood effect on mental health estimated with IPTW method.

TABLE 5.2: Estimated neighbourhood effect in terms of odds ratio (the reference is neighbourhood 6) without adjustment on observables (Crude), with the naive logistic regression (Logit) and with Inverse Probability of Treatment Weighting method (IPTW).

		Crude				**Logit**				**IPTW**			
Pop.	**Neigh.**	**Odds Ratio**	**Std.Err.**	**CI 95%**		**Odds Ratio**	**Std.Err.**	**CI 95%**		**Odds Ratio**	**Std.Err.**	**CI 95%**	
All	1	1.48	0.10	1.20	1.81	1.17	0.11	0.95	1.45	1.21	0.13	0.94	1.55
	2	1.31	0.10	1.09	1.59	1.34	0.10	1.11	1.62	1.33	0.11	1.07	1.64
	3	1.19	0.10	0.99	1.44	1.08	0.10	0.89	1.31	1.09	0.11	0.88	1.35
	4	1.30	0.10	1.07	1.59	1.15	0.10	0.94	1.41	1.15	0.11	0.92	1.43
	5	1.18	0.10	0.97	1.43	1.18	0.10	0.98	1.43	1.15	0.12	0.91	1.45
	7	1.46	0.10	1.19	1.78	1.33	0.10	1.09	1.63	1.33	0.11	1.06	1.66
	8	1.44	0.11	1.15	1.80	1.22	0.12	0.97	1.54	1.21	0.13	0.94	1.57
	9	1.29	0.10	1.05	1.58	1.25	0.11	1.02	1.54	1.25	0.12	1.00	1.58
	10	1.19	0.13	0.92	1.53	1.39	0.13	1.07	1.79	1.28	0.16	0.94	1.75
Women	1	1.49	0.12	1.19	1.87	1.22	0.12	0.97	1.55	1.32	0.14	1.00	1.75
	2	1.26	0.11	1.02	1.57	1.32	0.11	1.06	1.64	1.32	0.12	1.04	1.68
	3	1.23	0.11	1.00	1.52	1.14	0.11	0.92	1.41	1.18	0.12	0.93	1.49
	4	1.26	0.12	1.00	1.58	1.13	0.12	0.90	1.42	1.17	0.13	0.91	1.50
	5	1.11	0.11	0.89	1.38	1.13	0.11	0.91	1.41	1.14	0.14	0.87	1.48
	7	1.35	0.12	1.08	1.70	1.25	0.12	0.99	1.58	1.29	0.13	1.00	1.66
	8	1.45	0.13	1.13	1.86	1.27	0.13	0.98	1.64	1.27	0.15	0.95	1.69
	9	1.27	0.12	1.01	1.61	1.26	0.12	0.99	1.59	1.27	0.13	0.98	1.65
	10	1.26	0.15	0.95	1.68	1.54	0.15	1.15	2.05	1.43	0.17	1.02	2.00
Men	1	1.18	0.24	0.73	1.89	0.86	0.25	0.52	1.40	0.85	0.32	0.45	1.59
	2	1.46	0.20	0.99	2.18	1.40	0.20	0.95	2.10	1.37	0.24	0.86	2.18
	3	0.95	0.22	0.62	1.45	0.86	0.22	0.56	1.33	0.83	0.25	0.51	1.36
	4	1.32	0.22	0.86	2.03	1.17	0.22	0.76	1.81	1.07	0.25	0.65	1.75
	5	1.42	0.20	0.97	2.12	1.39	0.20	0.94	2.07	1.23	0.25	0.76	2.01
	7	1.72	0.21	1.15	2.62	1.59	0.21	1.05	2.42	1.50	0.24	0.93	2.42
	8	1.28	0.25	0.77	2.08	1.01	0.26	0.60	1.67	1.03	0.30	0.57	1.86
	9	1.31	0.22	0.85	2.02	1.23	0.22	0.80	1.92	1.23	0.26	0.74	2.04
	10	1.03	0.28	0.58	1.78	1.06	0.29	0.59	1.83	0.81	0.34	0.41	1.60

Table 5.3: Mental health standardized rate (STD Rate) and odds ratio for mental health for each neighbourhood with respect to the mean. Estimates obtained with IPTW as described in 4 (1) and with the inclusion of chronic conditions (2), significant effects (level 5%) are written in bold.

Neighbourhood	**Std Rate**	**Whole population**		**Female population**		**Male population**	
		(1)	(2)	(1)	(2)	(1)	(2)
1	10.67	0.942	0.953	0.945	0.958	0.929	0.935
2	10.75	0.984	0.986	0.979	0.980	0.985	0.990
3	11.11	**1.054**	**1.063**	**1.056**	**1.066**	1.048	1.057
4	11.24	**1.044**	**1.050**	1.032	1.039	1.063	1.066
5	9.88	**0.909**	**1.905**	**0.943**	**0.939**	**0.854**	**0.852**
6	10.64	0.966	0.956	0.995	0.982	**0.915**	**0.913**
7	10.92	1.032	1.031	1.025	1.023	1.044	1.044
8	11.62	1.042	1.045	1.000	1.002	**1.122**	**1.120**
9	11.04	1.019	1.017	1.025	1.022	1.016	1.015
10	11.64	1.018	1.006	1.007	0.996	1.054	1.039

Chapter 6

An original proposal: the MARMoT approach

6.1 Introduction

In this chapter we propose an original method to estimate neighbourhood effect on health of older people in Turin, a city in the North of Italy, adjusting for confounders. Our methodological proposal consists on a Matching procedure on Poset based Average Rank for Multiple Treatments (MARMoT). Poset theory is exploited to summarize individuals' confounders and the relative average rank is used to balance confounders and match individuals in many neighborhoods (treatments). This technique results to be particularly useful to balance confounders, even in frameworks in which the number of considered treatments is high.

In the last part of the chapter we estimate the neighbourhood effect on hospitalized fractures among older residents in Turin, a city located in the north of Italy. Our method allows to adjust for confounders, even when the number of treatments (neighbourhoods) is very high. Thus, main methodological contributions of this chapter consist in the definition of this new matching approach based on poset theory, its validation through a simulation study and its application to estimate the neighbourhood effect on hospitalized fractures among individuals aged 60 or more according to three different geographical partitions.

In the simulation and the empirical analysis we used data and variables described in chapter 3. In section 6.2 we describe the distribution of the confounders among the considered population in the 23 areas partition that is considered for the simulation study of this chapter. In section 6.3, after a brief introduction to poset theory, the original methodological proposal is explained in depth. In section 6.4 we describe the design structure and the results of the simulation study we performed in order to test the reliability of the original proposal. Section 6.5 illustrates the empirical application with real data comparing different neighbourhoods partitions.

6.2 Data and geographical partitions

As already explained in chapter 3, the city of Turin may be split in 10 districts, 23 areas or 94 zones, that are only partially hierarchical, for instance the same zone may be divided in two or more areas or districts.

Table 6.1 shows some summary statistics for the sizes of the populations in each geographical partition, considering the population selected in section 5.3.1. The ten districts have an average population of 22,583, with the least populated accounting for 10,608 individuals, and the most populated for 33,072. The populations of the areas range between 3,584 and 18,089, with a mean area population of 9,819. The number of individuals living in each zone varies even more. More extensive information about dimension and location of the considered zones, areas and districts are reported in table B.1, in appendix B.1.

Table 6.1: Distribution of the population size for each geographical partition.

Partition	*Minimum*	*1st Q.*	*Median*	*Mean*	*3rd Q.*	*Maximum*
10 Districts	10608	18777	21897	22583	29107	33072
23 Areas	3584	7976	9609	9819	12606	18089
94 Zones	3	625	1870	2402	3876	7758

In our empirical analysis, we estimate the neighbourhood effect on hospitalized fractures considering the three geographical partitions. In the case of the 94 zones, however, we needed to reduce the neighbourhoods considered because some of them were too small, as shown in the last row of table 6.1. We therefore excluded zones with a population of less than 625 (corresponding to the first quartile of the distribution of zone populations). The number of individuals living in the zones thus discarded account for only 3% of the whole sample population, and the final number of zones considered is 70. For the sake of brevity, in the simulation we focus on the intermediate partition, i.e. the city divided into 23 areas.

We reported the distribution of confounders in the 23 areas in table 6.2. It interesting to notice that there are a lot of differences among the 23 areas, such as the percentage of individuals with primary education or lower varies between 22% in area 3 and more

TABLE 6.2: Descriptive statistics on the outcome and confounders by neighbourhoods

	Areas																							
Variables	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**Total**
Hospitalized Fractures (%)	1.02	1.06	1.03	0.99	0.92	1.01	0.90	1.18	0.90	0.92	1.08	0.83	0.78	0.83	0.90	0.77	0.87	0.77	0.67	0.67	0.95	0.96	0.85	0.90
Female (%)	59.70	59.38	61.34	60.08	61.01	60.09	58.69	59.23	58.71	56.66	59.32	55.74	57.62	58.99	57.10	56.44	56.66	57.75	55.04	57.48	57.23	58.22	54.93	58.03
Age (Mean)	71.52	71.82	72.37	71.5	72.05	71.66	70.94	71.32	71.39	70.54	71.31	70.07	70.75	71.08	70.5	70.41	70.68	70.72	69.93	70.76	71.14	71.41	69.95	25.46
Educational attainment(%)																								
Primary or lower	31.27	35.41	22.58	44.18	38.65	39.02	55.11	46.34	49.68	46.87	37.99	42.47	44.18	46.55	62.81	59.70	59.07	60.15	62.84	61.40	38.28	24.07	62.52	46.94
Lower Secondary	25.55	29.52	25.72	35.09	30.56	30.25	30.88	30.51	33.24	33.63	33.40	34.35	32.68	32.48	27.87	30.34	29.98	29.28	28.97	27.67	29.83	26.44	28.31	30.69
Upper Secondary or higher	43.18	35.07	51.70	20.73	30.79	30.73	14.01	23.14	17.09	19.50	28.62	23.18	23.14	20.97	9.33	9.96	10.95	10.57	8.19	10.94	31.89	49.50	9.17	22.36
Family composition (number of components)(%)																								
Alone (1)	37.66	33.91	34.30	32.06	33.74	32.79	32.52	30.90	30.60	25.41	29.44	23.89	27.78	29.60	24.54	26.84	27.43	29.93	21.70	28.06	29.85	29.16	20.03	28.73
Married couple (2)	30.99	36.61	35.95	42.22	38.37	39.40	39.98	41.77	43.04	44.00	43.54	46.30	44.23	42.97	43.32	46.10	45.92	42.60	45.50	42.23	41.32	40.38	45.59	42.28
Married couple (>3)	17.82	17.47	17.13	15.45	15.55	16.41	16.46	16.96	16.72	20.44	17.17	20.72	18.62	17.25	19.58	18.11	16.62	17.61	22.35	18.15	19.39	19.58	23.92	18.41
No married couple (>2)	13.52	12.02	12.62	10.27	12.35	11.39	11.04	10.37	9.64	10.15	9.85	9.09	9.37	10.19	12.56	8.95	10.03	9.86	10.45	11.57	9.43	10.88	10.45	10.58
Last observed professional condition(%)																								
No observed work	14.70	14.67	13.21	13.11	14.22	15.14	17.85	16.19	16.78	13.43	12.21	11.22	12.65	13.18	14.25	13.44	16.79	15.42	14.35	18.94	15.15	12.64	11.90	14.00
Home-maker	31.61	33.77	36.00	36.97	35.83	34.41	32.07	34.80	32.96	34.00	35.57	34.86	35.93	35.18	35.53	34.80	34.05	33.49	34.26	33.22	33.59	36.71	36.52	34.74
Entrepreneur	16.37	10.57	17.33	4.95	8.44	8.31	3.57	6.06	5.04	4.69	6.73	5.31	6.48	6.06	2.17	2.75	2.44	2.46	2.10	2.54	11.75	19.27	1.73	6.34
White collars	24.22	26.04	24.27	24.78	25.13	25.05	20.59	23.80	20.73	24.33	26.87	26.52	24.45	24.45	15.98	18.16	18.85	18.66	15.57	17.02	25.08	22.41	14.97	22.33
Manual Workers	13.11	14.95	9.19	20.18	16.39	17.09	25.92	19.15	24.49	23.55	18.62	22.09	20.49	21.13	32.07	30.85	27.87	29.97	33.72	28.28	14.43	8.97	34.89	22.59
Region of Birth(%)																								
Piedmont	51.22	54.01	60.27	51.74	52.09	50.99	39.18	50.82	50.50	45.49	52.10	46.21	49.15	48.56	29.75	38.01	40.26	37.73	29.30	39.70	64.84	69.60	29.95	45.93
North of Italy	13.38	12.88	14.29	14.25	13.32	13.39	11.83	12.45	14.75	14.82	14.89	14.60	16.40	17.52	12.08	15.07	13.73	12.70	12.32	15.37	15.07	14.34	12.70	14.12
Center of Italy	3.85	3.30	3.60	2.61	2.53	2.79	2.95	2.41	2.60	3.63	3.39	3.68	2.93	2.95	1.82	3.09	3.52	2.85	2.53	2.86	2.15	3.14	2.35	2.97
South of Italy	26.40	25.28	17.34	26.45	26.60	28.33	40.54	29.56	24.12	29.93	24.96	29.58	26.33	25.81	43.61	37.40	36.56	40.18	46.98	35.90	14.73	9.10	48.57	30.93
Outside of Italy	5.14	4.53	4.50	4.96	5.45	4.50	5.50	4.75	8.03	6.11	4.67	5.94	5.20	5.16	12.74	6.44	5.93	6.55	8.88	6.17	3.21	3.82	6.43	6.05
Home owner (%)	65.80	77.01	74.58	77.98	73.76	75.15	74.93	76.74	82.48	78.70	80.46	82.25	80.42	75.86	66.24	74.36	76.94	78.17	72.04	61.75	81.47	82.89	79.19	76.34

than 62% in areas 15, 19 and 23. The first row of table 6.2 reports the percentage of hospitalized fractures in 2002 in each district. The selected outcome affects a small portion of population, 0.9% of elderly living in Turin, with some differences among neighbourhoods: from district 19 that presents the lowest proportion (0.67%) to the district 8 with 1.18% of hospitalized fractures in 2002.

6.3 Matching on Poset based Average Rank for Multiple Treatment (MARMoT)

6.3.1 Introduction to poset theory

Partially ordered set (poset) is, in mathematics, a set of elements where a binary relation that indicates an order can be traced, the word "partial" refers to the fact that not every pair of elements needs to be comparable. Poset theory is a theoretical field between graph theory and discrete mathematics that quickly developed after the 1970s thanks to technological advances that made greater computational efforts manageable (Brüggemann and Patil, 2011). The main concepts needed to understand why this method is useful to overcome the curse of dimensionality without using a parametric model or introducing some subjective criteria are explained with a toy example.

When dealing with a population, the people comprising it can be ranked and ordered using a single variable: level of education, for instance, enables two different individuals to be arranged in an order. From the mathematical standpoint, an order is a binary relation between the elements in a set that respects specific properties. Let P be a set, an order on P is a relation ($\leq$) between two elements in the set P such that, for all $x, y, z \in P$, the following properties hold:

- Reflexivity: $x \leq x$
- Antisymmetry: $x \leq y$ and $x \geq y$ implies $x = y$
- Transitivity: $x \leq y$ and $y \leq z$ implies $x \leq z$.

A set equipped with such a relation is said to be ordered. If the comparison is drawn using several variables, it may be that some elements are neither equal nor ordered, in which case they are defined as incomparable (Davey and Priestley, 2002). The word "partially" is added to "ordered set" when some of its elements are incomparable, so the order relation has to be changed to a partial order relation, which takes the incomparability (indicated with $||$) of the elements into account:

TABLE 6.3: Toy example for a group of observations.

Subject	*Age*	*Education*	*Homeowner*
A	0	0	0
B	1	0	0
C	0	1	0
E	1	1	0
G	0	1	1
H	1	1	1

Incomparability: $x||y \leftrightarrow x \nleq y \quad and \quad y \nleq x, \quad x, y \in P$.

Comparing the individuals in a population gives rise to a list of comparabilities and incomparabilities, which can be represented in a graphic form called a Hasse diagram. This diagram represents the elements in a poset: each node is an element, two or more equal elements still form one node, and every line segment is an order relation between comparable objects. Let us suppose that we have a population comprising six individuals characterized by three dichotomous variables, as represented in 6.3: age (which takes a value of 0 for individuals who are between 60 and 70 years old, and 1 if they are older); education (which takes a value of 0 if they have a higher education, and 1 otherwise); and homeowner (which takes a value of 0 if they own the house in which they live, and 1 otherwise). These variables are ordered according to the risk of experiencing the outcome, where a value of 1 corresponds to the highest risk of hospitalized fracture. In this example, for the sake of simplicity, we included only dichotomous variables, but categorical and continuous variables may be also considered in a poset. However, in order to contain the entropy of the poset, it is recommended to reduce continuous variables in meaningful classes. Indeed, since the AR is just used as a balancing tool, the most suitable classes are those that help providing a better balance of the continuous variable among different treatment groups and that may guarantee the smaller distortion of results.

A Hasse diagram can be used to visualize the order relations between the elements in a poset, and it is based entirely on the order of the elements, disregarding any quantitative information.

In Figure 6.1(a), the six individuals are represented by their profile in the Hasse diagram, where each node stands for a profile. When two individuals are comparable, they are connected by line segments in the diagram, like A and B or B and E, whereas there is no ascending or descending path between incomparable elements, like B and C.

The list of all the ranks that each individual may occupy is shown in part (b) of Figure 6.1, where all the linear extensions of the poset are listed. Linear extensions are

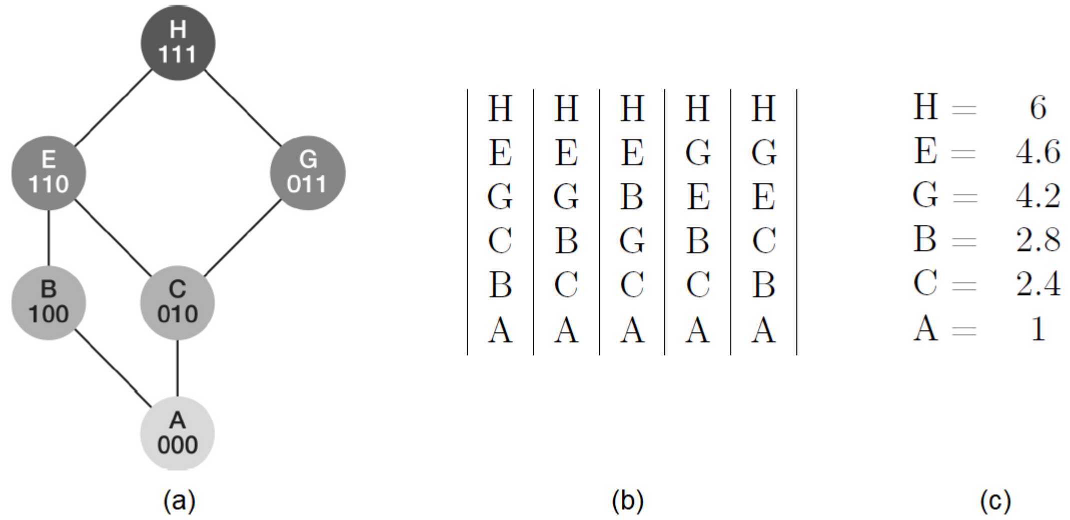

FIGURE 6.1: A poset and its linear extensions: part (a) represents the Hasse diagram of subjects in table 6.3; part (b) lists all linear extensions for subjects in table 6.3; and in part (c) the exact average rank for individuals of table 6.3 is showed.

all the possible rankings of elements in the poset that respect its comparabilities (the connections in the Hasse diagram) and incomparabilities (Brüggemann and Patil, 2011; Davey and Priestley, 2002). The average rank (AR) of a node represents the mean of all the ranks that the element occupies in all possible linear extensions, starting from the known order relations, as listed in Figure 6.1 part (c).

The AR is a single value for each element in the set that describes the relative position of a given element with respect to the rest of the population. It can be normalized in the interval [0;1]. Even if the AR has been used also as a composite indicator to represent complex or latent concepts, its involvement in the MARMoT approach is just as a balancing tool. Since the AR purpose is to reduce data dimensionality and balance on observable individuals' characteristics, there is no need in finding a substantial interpretation to AR values.

6.3.2 Approximating the average rank

If the number of individuals and variables increases, the linear extensions become too many to be examined thoroughly, and it becomes computationally almost impossible to find the exact AR as in the example in Table 6.3. That said, satisfactory approximations of the number of linear extensions of a poset can be found in works by Dyer *et al.* (1991), and De Loof (2009).

Researchers have used two main approaches to obtain a computationally efficient calculation of the AR, by sampling linear extensions (Fattore, 2016; Lerche and Sorensen, 2003), or defining an approximation formula. Different approximation formulas have been proposed in the literature, such as the Local Partial Order Model (Brüggemann and Carlsen, 2011), or the one based on Mutual Probabilities (De Loof, 2009). The present work is based on De Loof's approach (2009) because it provides better results than other methods in terms of accuracy with a large sample size (De Loof *et al.*, 2011).

Two concepts help us to understand this approximation, for a sample P with $|P|$ elements:

The rank probability $P(rank(x) = i)$ is the fraction of linear extensions in which an element's rank equals i, where $i = 1, \ldots, |P|$ all possible ranks in the sample of size $|P|$.

The mutual rank probability $P(x > y)$ of two elements $x, y \in P$ is the fraction of linear extensions in which the element x is ranked higher than element y.

Now we can establish a relation between the last-mentioned two concepts and the real AR of elements x, $\bar{h}(x)$, starting from a sample P with $|P|$ elements, including x and y:

$$\bar{h}(x) = \sum_{i=1}^{|P|} i \cdot P(rank(x) = i) = 1 + \sum_{y=1}^{|P|} P(x > y). \tag{6.1}$$

In other words, the first part of formula 6.1 describes the real AR value, $\bar{h}(x)$, as the expected value, multiplying each possible rank value i by the fraction of linear extensions in which the element's rank equals i. The second part of formula 6.1 expresses the real AR value as the sum of all the mutual rank probabilities that involve the element x. Starting from this formula, we need to find an approximation for the mutual rank probability. To do so, we have to define three subsets of the poset P, given a generic element $x \in P$:

Downset: $O(x) = \{y \in P : y \leq x\}$;

Upset: $F(x) = \{y \in P : y \geq x\}$;

Incomparables: $U(x) = \{y \in P : y || x\}$

If $y \in O(x)$, then $P(rank(x) > rank(y))$ equals 1, and if $y \in F(x)$, then $P(rank(x) > rank(y))$ equals 0, so the mutual rank probabilities only need to be approximated with respect to the reciprocal ranks of the incomparable elements. The following approximation was proposed by Brüggemann *et al.* (2004)

$$P^*(x > y) = \frac{[o(x) + 1][f(y) + 1]}{[o(x) + 1][f(y) + 1] + [o(y) + 1][f(x) + 1]}, \tag{6.2}$$

TABLE 6.4: A numerical example of the approximation of the average rank according to De Loof (2009) approach.

x	$o(x)$	$f(x)$	$U(x)$	Pr^* $(x > y)$						$\tilde{o}(x)$	$\tilde{f}(x)$	AR(x)
				$y = A$	$y = B$	$y = C$	$y = E$	$y = G$	$y = H$			
A	0	5	0	.	0.20	0.25	0.08	0.10	0.03	0.00	5.00	1.00
B	1	2	C, G	0.80	.	0.57	0.25	0.31	0.10	1.88	3.12	2.90
C	1	3	B	0.75	0.43	.	0.20	0.25	0.08	1.43	3.57	2.43
E	3	1	G	0.92	0.75	0.80	.	0.57	0.25	3.57	1.43	4.57
G	2	1	B, E	0.90	0.69	0.75	0.43	.	0.20	3.12	1.88	4.10
H	5	0	0	0.97	0.90	0.92	0.75	0.80	.	5.00	0.00	6.00

where $o(x) = |O(x) \setminus \{x\}|$ and $f(x) = |F(x) \setminus \{x\}|$ are respectively the number of elements in the downset and the upset of x without $\{x\}$. Two more quantities are needed to approximate the AR according to the De Loof (2009) formula, $\tilde{o}(x)$ and $\tilde{f}(x)$:

$$\tilde{o}(x) = o(x) + \sum_{y \in U(x)} P^*(x > y) \quad \textit{and} \tag{6.3}$$

$$\tilde{f}(x) = f(x) + \sum_{y \in U(x)} P^*(x < y), \tag{6.4}$$

and the AR approximation proposed by De Loof (2009) is

$$AR(x) = o(x) + 1 + \sum_{y \in U(x)} \frac{[\tilde{o}(x) + 1][\tilde{f}(y) + 1]}{[\tilde{o}(x) + 1][\tilde{f}(y) + 1] + [\tilde{o}(y) + 1][\tilde{f}(x) + 1]}. \tag{6.5}$$

In other words, using formula 6.5, the AR of x is given by the number of elements in its downset and the sum of probabilities of being a part of x's downset for all incomparable elements with respect to x, using the approximation of the *mutual rank probabilities.* Following the toy example in Table 6.3, the steps needed to approximate the AR with the De Loof (2009) approach are solved in Table 6.4, including the estimation of the AR.

In the present work, the approximated AR was computed using the R software, with an R function proposed by Caperna (2016) that can cope with large datasets (Boccuzzo and Caperna, 2017; Caperna and Boccuzzo, 2018).

6.3.3 The Matching

We use our MARMoT technique to address the so-called curse of dimensionality, the need to summarize confounders, applying a poset-based AR of the individuals. The individuals' characteristics are summarized by unique numbers, and individuals who have a similar AR have similar profiles. Using this score, which summarizes individuals' characteristics, enables us to proceed with a matching whereby each individual in a given neighbourhood is allocated an individual with a similar AR in all the other neighbourhoods, and those who cannot be matched are discarded in order to respect the overlap condition and make all neighbourhoods comparable simultaneously, without any need for a template.

Once the AR has been computed, the first step is to build a frequency table where each row corresponds to one value of the AR (AR_r, $r = 1, \dots, R$), and each column represents a treatment group (t, $t = 1, \dots, K$).

In order for each value of the AR to be represented equally in all the treatment groups, the desired result would be a table where $f_{r,1} = f_{r,2} = \cdots = f_{r,t} = \cdots = f_{r,K} = f_r$, $\forall t = 1, \dots, K$ in every row r.

Thus, for every row, we must choose the most appropriate frequency f_r for each AR value to impose in the balanced population. In the artificial final population, the distribution of AR values will be balanced in all the treatments groups so as to balance all confounders too. At the end of the matching procedure, each AR_r value will be present in the balanced population $K * f_r$ times, with f_r individuals in each of the K treatment groups. The value for f_r may be chosen according to different criteria: for example, it may be the maximum, the mean, the median or the minimum of the frequencies in row r. In this work, we define the reference f_r as

$$f_r = \begin{cases} 1 & if \quad median(f_{r,1}, f_{r,2}, \dots, f_{r,K}) = 0 \\ median(f_{r,1}, f_{r,2}, \dots, f_{r,K}) & otherwise. \end{cases} \tag{6.6}$$

Instead of discarding all the AR values with $median(f_{r,1}, f_{r,2}, \dots, f_{r,K}) = 0$, we set the minimum value of f_r at 1 in order to have a matched population that includes all the profiles in the real population. The choice of the value for f_r may affect both the final dimension of the balanced dataset, and the performance of the MARMoT method in terms of balance. For instance, if we define f_r as the maximum of the frequencies in row r, the final dimension of the dataset will be more than double the dimension obtained with the previous definition and also the quality of matches will be worse. Indeed, for AR values where the frequency matrix is sparse, individuals are duplicated creating distortion and noise in the final dataset.

Having established the frequency that each value of AR should have in each treatment, the algorithm could proceed in three different ways, depending on the dimensions of $f_{r,t}$ and f_r, for every r and every t:

1. if $f_{r,t} = f_r$: all individuals with AR_r in the treatment group t are copied in the final dataset.

2. if $f_{r,t} \neq f_r$ and $f_{r,t} \neq 0$: a random sample with replacement of size f_r is selected from among the individuals with AR_r in the treatment group t, and included in the final dataset.

3. if $f_{r,t} = 0$: a random sample with replacement of size f_r is selected from among the individuals with an AR close enough (with a given tolerance) to AR_r in treatment group t, and included in the final dataset. If there are no individuals close enough, then all individuals with an AR equal to AR_r have to be deleted from the final dataset.

While points (a) and (b) are just a matter of matching individuals with identical AR values, point (c) is the trickiest because it involves inexact matching, and possibly excluding some individuals from the final dataset. In this work we define the tolerance interval as $[AR_r - \frac{S_{AR}}{4}; AR_r + \frac{S_{AR}}{4}]$, considering as a caliper the value $\frac{S_{AR}}{4}$ that is recommended in the propensity score matching literature (Cochran and Rubin, 1973; Lunt, 2013). Thus, if all frequencies $f_{.,t}$ that correspond to AR values included in the interval $[AR_r - \frac{S_{AR}}{4}; AR_r + \frac{S_{AR}}{4}]$ equal 0, the AR_r value will not be considered in any treatment group for the final population, as if its row in the frequency table had never existed. This criterion ensures that the overlap assumption is respected.

Once the MARMoT algorithm has matched the individuals and balanced the confounders, any common causal inference estimand can be used to calculate the effect of a treatment. In the following paragraphs, we use the ATT to estimate the neighbourhood effect.

6.4 Simulation study

Before using the MARMoT method to estimate the neighbourhood effect on real data, we tested it with some simulations in two different scenarios for allocating individuals to 23 treatments, and two for the occurrence of the outcome.

6.4.1 Simulation design

To keep our simulation close to the real situation of interest, we considered the real population of Turin and the individuals' observed characteristics. Starting from the seven confounders described in section 3.2.1, we just simulated the treatment allocation and the occurrence of the outcome. Since the computation of the AR depends only on individual variables (which come from the observed population and are not simulated artificially), and not on the treatment or the outcome, AR values computed directly on the observed data could be used, meaning that they were based exclusively on the real population, not on simulated values.

We chose to simulate the treatment allocation and the occurrence of the outcome according to two different scenarios, their combination giving rise to four scenarios in all.

In the first scenario, the treatment allocation equation is simple and close to the real situation. The treatment is generated through a multinomial logistic model, taking neighbourhood 20 (the one with the lowest crude hospitalized fractures rate) for reference. Thus, for each neighbourhood t, and each individual i, the treatment equation is

$$\begin{aligned} ln\left(\frac{Pr(T_i = t)}{Pr(T_i = 20)}\right) &= \beta_0^t + \beta_1^t * Gender_i + \beta_2^t * LowerSecondary_i + \\ &+ \beta_3^t * UpperSecondary_i + \beta_4^t * Age65-69_i + \beta_5^t * Age70-74_i + \\ &+ \beta_6^t * Age75-79_i + \beta_7^t * Age > 79_i + \beta_8^t * MarriedCouple(2)_i + \\ &+ \beta_9^t * MarriedCouple(>3)_i + \beta_{10}^t * NoMarriedCouple(>2)_i + \\ &+ \beta_{11}^t * HomeMaker_i + \beta_{12}^t * Entrepreneur_i + \beta_{13}^t * WhiteCollars_i + \\ &+ \beta_{14}^t * Manualworkers_i + \beta_{15}^t * NorthofItaly_i + \\ &+ \beta_{16}^t * CenterofItaly_i + \beta_{17}^t * SouthofItaly_i + \\ &+ \beta_{18}^t * OutsideofItaly_i + \beta_{19}^t * Homeowner_i. \end{aligned} \tag{6.7}$$

In order to choose values for the coefficients, we estimated a multinomial logistic model on the whole population. The result was a matrix with 23 rows and 20 columns containing all coefficients β_v^t for $t = 1, ..., 19, 21, 22, 23$, and coefficients β_v $v = 0, ...19$ for the other variables in the model. These coefficients were perturbed by adding a random value coming from a uniform distribution between -0.01 and $+0.01$, and rounded up or down to just three decimals.

The second scenario envisages a more complex treatment allocation equation, which includes all the interactions between the seven variables considered. As in the first scenario, the choice of parameters for these treatment allocation equations was based on those estimated by a multinomial logistic model, perturbed by a uniform distribution between -0.1 and $+0.1$, and rounded up or down to just three decimals.

As for generating the outcome, it was simulated for two different scenarios, one simple and one that also included significant interaction terms.

To simulate the outcome, we used a logistic regression model and the equation included dummy variables representing the neighbourhood effect on the outcome.

As in the generation of the treatment, the choice of parameters for generating the outcome equation was based on those estimated by the logistic regression model, perturbed by a uniform distribution between -0.1 and $+0.1$, and rounded up or down to just three decimals.

Only statistically significant (at 10% level) interactions between variables were introduced in the second outcome scenario, and the path for selecting the values of the coefficients in the model was the same as for the first outcome scenario.

For the simulation of the outcome we used a logistic regression model and the equation includes also dummy variables that represent the neighbourhood effect on the outcome. The first outcome scenario is simulated with:

$$\begin{aligned} ln\left(\frac{Pr(Y_i=1)}{Pr(Y_i=0)}\right) &= \beta_0 + \beta_1 * Gender_i + \beta_2 * LowerSecondary_i + \\ &+ \beta_3 * UpperSecondary_i + \beta_4 * Age65-69_i + \\ &+ \beta_5 * Age70-74_i + \beta_6 * Age75-79_i + \beta_7 * Age>79_i + \\ &+ \beta_8 * MarriedCouple(2)_i + \beta_9 * MarriedCouple(>3)_i + \\ &+ \beta_{10} * NoMarriedCouple(>2)_i + \beta_{11} * HomeMaker_i + \\ &+ \beta_{12} * Entrepreneur_i + \beta_{13} * WhiteCollars_i + \\ &+ \beta_{14} * Manualworkers_i + \beta_{15} * NorthofItaly_i + \\ &+ \beta_{16} * CenterofItaly_i + \beta_{17} * SouthofItaly_i + \\ &+ \beta_{18} * OutsideofItaly_i + \beta_{19} * Homeowner_i + \\ &+ +\beta_{20} * Neigh1_i + \beta_{21} * Neigh2_i + \beta_{22} * Neigh3_i + \ldots \\ &+ \beta_{38} * Neigh19_i + \beta_{39} * Neigh21_i + \beta_{40} * Neigh22_i + \\ &+ \beta_{41} * Neigh23_i, \end{aligned} \tag{6.8}$$

where Neigh1, Neigh2, ... , Neigh19, Neigh21, Neigh22, Neigh23 are dichotomous variables that take value 1 if the individual i lives in the considered neighborhood and 0 otherwise. The reference, as before, is the neighborhood number 20. As in treatment generation, parameters for the outcome generation equation were chosen according to those estimated by the logistic regression model, perturbed by a uniform distribution between -0.1 and $+0.1$ and rounded to have just three decimals.

In the second outcome scenario only statistically significant (at 10% level) interactions between variables were introduced and the path to select values of coefficients of the model is the same to those described for the first outcome scenario.

6.4.2 Results

The main results of the simulations are shown in Table 6.5, where column T indicates the above-described treatment allocation scenarios (coded as 1 for the linear and additive, and 2 for the one with interactions), and column O indicates the outcome occurrence scenarios (coded as 1 for the linear and additive, and 2 for the one with interactions). The first part of Table 6.5 shows the results of the simulation as described in the previous section, the differences between the scenarios, the differences in the distribution of the individuals among the neighbourhoods, and the distribution of the outcomes among the neighbourhoods.

We examined the initial balance of the two scenarios in all 1000 simulations using the ASB. Having 23 neighbourhoods and seven variables (for a total of 24 levels), we chose to summarize the information by computing the minimum, the 1^{st} quartile, the mean, the median, the 3^{rd} quartile and the maximum of all the ASB, counting ASB values over 5% and 10% for each iteration. The means of these values among all 1000 simulations before and after the balancing procedure for each scenario are given in Table 6.6. The balance was much improved in both scenarios after the matching procedure, which fixed even extremely unbalanced situations. After matching, the mean number of ASB over 10% corresponded to one tenth of the number beforehand. The central part of Table 6.5 shows the means (among the simulations) of the number of ASB higher than 5% and 10% before and after adjusting for each neighbourhood. From these results, we can see that our MARMoT method greatly improves the balance of confounders among neighbourhoods: it achieves a five-fold reduction in the number of ABS over 5%, and an almost ten-fold reduction in those over 10%, in both the treatment scenarios.

The last part of table 6.5 shows the mean bias of the estimates with respect to the unbalanced and balanced situations. It is expressed in terms of difference between the estimates of ATT and the true ATT value, then since those differences were really small, we multiplied them for 100 in order to read results easily. As already said, the bias is really small, lower than 0.003, in all treatments in all the four considered scenarios. Higher values are presented in the second treatment scenario, in particular in correspondence of neighbourhood 8, 11, 20 and 23 in both the outcome scenarios and in neighbourhood 12 in the first outcome scenario.

TABLE 6.5: Results of simulations in the two treatment (T) and outcome (O) scenarios: (a) mean percentage of the distribution of individuals among neighbourhoods in the 1000 iterations; (b) mean percentage of the occurrence of outcomes among neighbourhoods in the 1000 iterations; (c) mean of the number of ASB greater than 5% before adjustment for each neighbourhood; (d) mean of the number of ASB greater than 5% after adjustment for each neighbourhood; (e) mean of the number of ASB greater than 10% before adjustment for each neighbourhood; (f) mean of the number of ASB greater than 10% after adjustment for each neighbourhood.

	T	O	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
			Simulation design																						
(a)		.	3.49	3.98	4.28	2.86	4.31	4.65	3.54	3.93	3.59	5.91	5.70	7.76	7.83	5.47	5.83	5.71	1.92	4.63	4.25	2.24	1.59	2.13	4.39
(b)	1	1	1.02	1.14	0.05	0.50	0.68	0.99	0.79	3.75	0.49	0.47	2.96	1.76	0.01	0.12	2.46	0.42	0.60	0.41	0.16	0.00	0.51	0.46	1.64
(b)		2	1.97	1.21	0.27	0.75	0.79	0.97	1.04	3.48	0.30	0.75	3.04	0.53	0.47	0.63	1.86	0.15	0.96	0.19	0.13	0.04	0.77	0.80	1.76
(a)		.	4.11	3.58	4.03	3.54	4.18	4.84	4.02	4.44	4.08	4.54	6.04	7.44	6.15	5.61	6.35	6.30	1.83	3.82	4.63	2.18	1.93	1.86	4.50
(b)	2	1	0.81	1.06	0.04	0.38	0.49	1.00	0.65	3.16	0.61	0.64	2.70	1.92	0.01	0.15	2.44	0.50	0.60	0.57	0.22	0.00	0.49	0.59	1.34
(b)		2	1.93	1.13	0.20	0.69	0.67	1.01	0.92	3.09	0.43	1.04	2.68	0.59	0.61	0.68	1.80	0.15	0.95	0.20	0.11	0.04	0.77	0.86	1.47
			Balance before and after matching																						
(c)	1	.	15.4	12.6	15.8	10.2	15.9	11.2	11.2	5.1	9.6	6.3	9.7	11.3	5.7	3.5	14.8	10.1	10.2	7.6	16.8	10.3	10.3	11.5	16.5
(d)	1	.	4.3	1.5	4.0	0.2	0.8	0.6	0.8	0.0	1.1	0.1	0.1	0.3	0.0	0.0	3.2	1.4	3.7	2.3	4.5	4.5	6.7	8.6	3.8
(e)	1	.	9.8	8.0	10.1	1.6	5.8	3.7	5.8	0.5	3.2	0.0	5.7	2.8	0.9	1.0	9.3	6.6	6.0	6.2	10.4	8.6	7.9	7.9	10.0
(f)	1	.	1.2	0.0	2.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.0	1.6	0.0	0.7	0.8	2.3	5.0	0.5
(c)	2	.	15.3	13.8	14.8	14.0	16.9	13.8	12.5	7.6	16.8	10.8	9.8	14.9	8.1	7.1	15.9	10.0	13.5	12.6	14.9	10.0	16.4	11.7	16.2
(d)	2	.	5.1	1.9	4.6	0.1	1.1	1.0	0.5	0.0	1.4	0.0	0.8	0.0	0.0	0.0	2.2	1.2	5.1	3.9	5.1	5.1	5.8	9.2	5.5
(e)	2	.	12.0	9.0	10.0	7.0	9.1	7.6	6.1	1.4	8.4	2.1	5.1	8.4	1.0	0.3	11.8	6.1	8.1	8.6	11.0	8.1	11.8	8.8	12.8
(f)	2	.	1.5	0.0	2.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	0.0	2.5	0.5	1.0	1.1	2.0	5.7	0.3
			Mean of the ATT bias after adjustment : (ATT-TRUEATT)*100																						
	1	1	0.31	0.12	0.01	−0.13	0.13	0.12	0.20	0.20	0.20	−0.11	−0.20	−0.34	−0.19	0.04	0.09	−0.07	0.13	0.30	−0.08	0.56	−0.07	−0.10	−0.38
	1	2	0.50	0.21	0.19	−0.19	0.22	0.20	0.11	0.24	−0.03	−0.14	−0.17	−0.35	−0.19	0.02	0.23	−0.15	0.06	0.11	−0.13	0.45	−0.18	−0.01	−0.40
	2	1	0.81	0.12	1.03	0.33	0.65	0.30	0.58	−2.16	0.70	0.29	−2.71	−2.47	0.86	0.94	−1.18	0.47	0.53	0.93	0.81	1.60	0.39	0.31	−2.38
	2	2	0.51	0.27	1.01	0.01	0.66	0.48	0.20	−1.96	0.71	−0.13	−2.58	−0.30	0.18	0.35	−0.29	0.70	0.00	0.94	0.80	1.47	−0.10	0.10	−2.41

Table 6.6: Mean of ASB summary statistics in the first and the second scenario before ad after balance among 1000 simulations.

Scenario	*Balance*	*Min*	1^{st} *Quartile*	Median	3^{rd} Quartile	Max	Mean	Over 5%	Over 10%
First	Before	0.01	1.98	4.43	9.59	63.72	8.01	252	132
	After	0	0.58	1.3	2.59	16.95	2.12	52	15
Second	Before	0.02	2.42	5.63	11.89	68.41	9.1	297	175
	After	0	0.62	1.37	2.71	16.84	2.24	60	18

Table 6.7: Mean of ASB summary statistics in the empirical study in different geographical partitions before ad after balance.

Partitions	*Balance*	*Min*	1^{st} *Quartile*	Median	3^{rd} Quartile	Max	Mean	Over 5%	Over 10%
10	Before	0.012	1.689	3.563	7.587	56.207	7.242	101	46
Districts	After	0	0.192	0.427	0.937	8.948	0.846	5	0
23	Before	0.072	1.934	4.198	9.774	63.763	7.938	248	132
Areas	After	0.003	0.482	1.169	2.33	15.802	1.973	51	11
70	Before	0.008	2.556	5.723	12.287	105.132	10.02	914	522
Zones	After	0.008	1.539	3.523	7.075	55.625	5.725	624	265

6.5 Empirical Results

n this section, we use our MARMoT technique to estimate neighbourhood effects considering 10 districts, 23 smaller areas and 94 zones. As explained in section 6.2, rather than considering all 94 zones, we selected 70 of them with a sufficient number of individuals (more than 625) to avoid the individuals in the excluded neighbourhoods causing problems in the balancing procedure. Table 6.7 shows that the MARMoT method substantially reduces the ASB in the three partitions, but slightly less successfully in the case of the 70 zones.

The computational time required for the MARMoT is acceptable, as the procedure to balance the 10 districts took less than 18 minutes, the one for the 23 areas took 36 minutes, and the balancing of the 70 zones took 116 minutes.

Figures 6.2, 6.3 and 6.4 plot the neighbourhood effects estimated in the three geographical partitions, expressed as $ATT * 100$, before and after the MARMoT procedure. As expected, there are several differences in the ATT values before and after matching. Considering smaller areas also enabled us to identify the less smoothed neighbourhood effects, even though it proved necessary to discard an extensive portion of the 70-zone partition (grey in figure 6.4). These discarded areas are scarcely populated, however: the eastern part of the map (in grey) is hilly and essentially very different and scarcely comparable with the rest of Turin. More extensive information about the estimated neighbourhood effect in the considered zones, areas and districts are reported in table B.2, in appendix B.2.

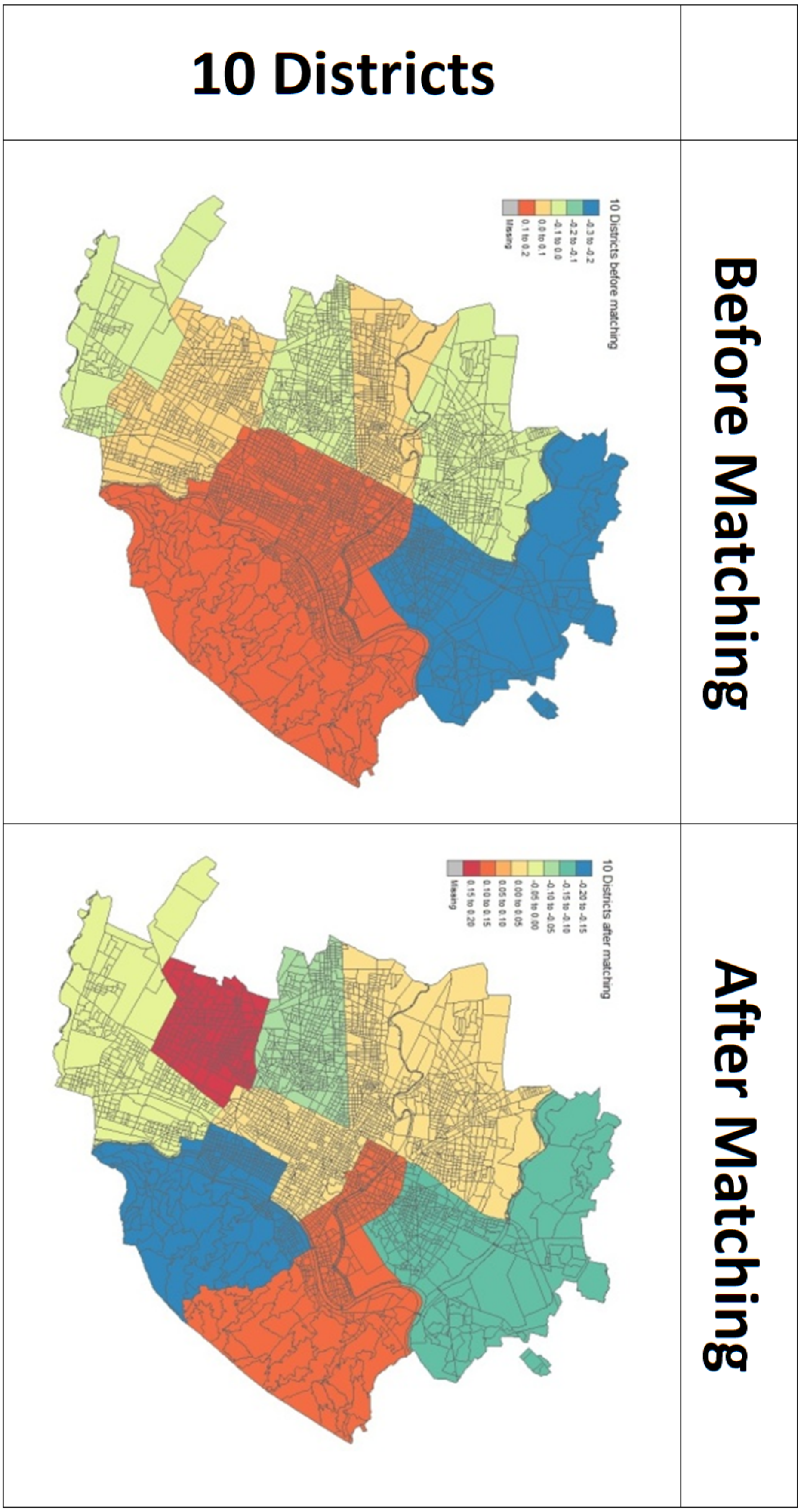

FIGURE 6.2: Neighbourhood effect estimates (ATT) before and after MARMoT in the 10 districts geographical partition.

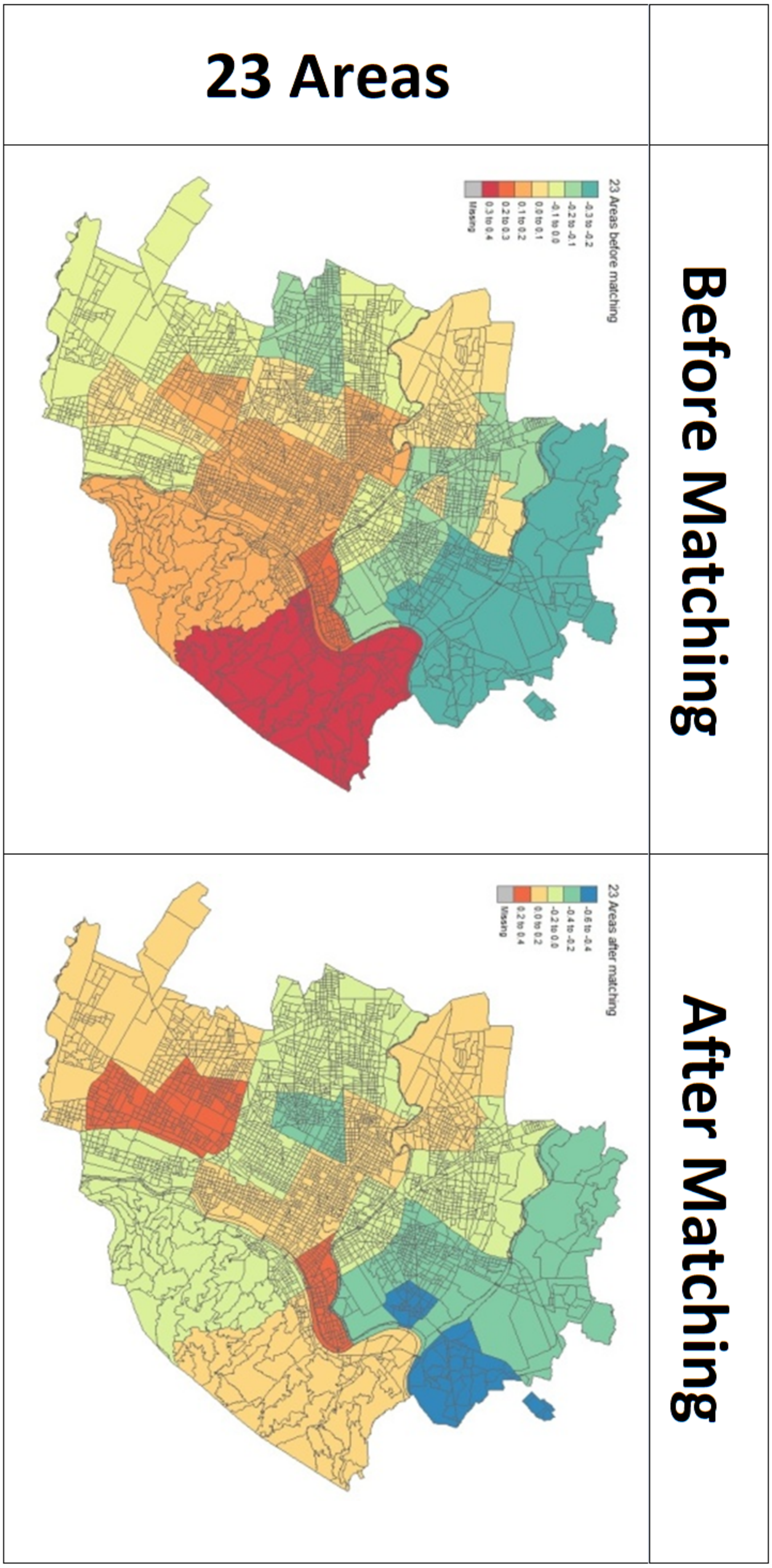

FIGURE 6.3: Neighbourhood effect estimates (ATT) before and after MARMoT in the 23 areas geographical partition.

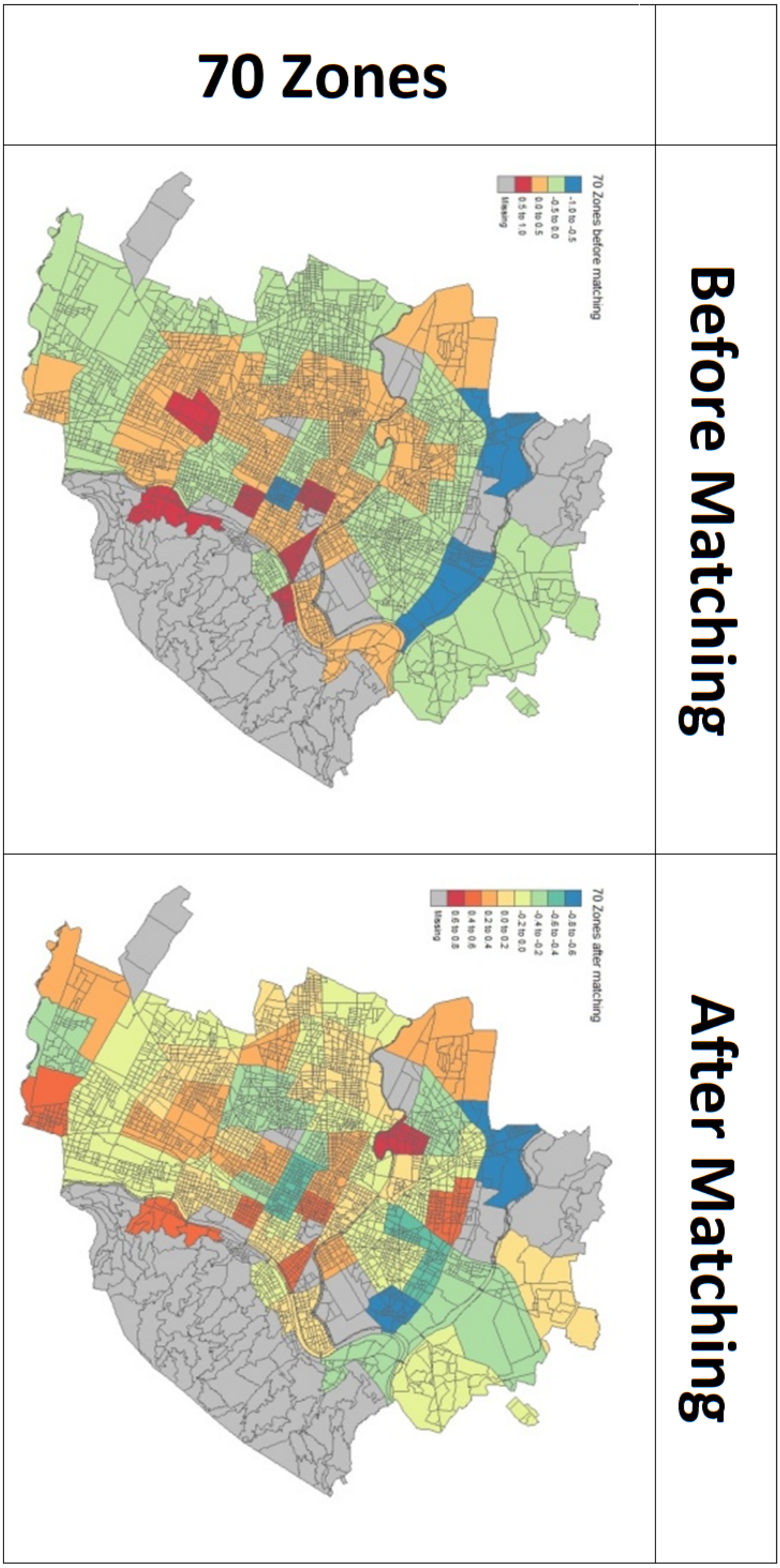

FIGURE 6.4: Neighbourhood effect estimates (ATT) before and after MARMoT in the 70 zones geographical partition.

6.6 Conclusions

The aim of this chapter was to estimate the neighbourhood effect on hospitalized fractures involving Turin residents over 60 years old, using a propensity score matching approach to adjust for confounders. To this end, we adopted an original approach based on poset theory, that we labelled MARMoT. The main idea behind our method was to obtain a population in which each profile, each combination of confounders summarized by the poset-based AR, was equally represented in all the treatment groups. The MARMoT approach proved very useful in balancing for confounders and reducing biases in our estimates. The matching involved is not bound to subjective choices (of the template, for instance), and the computation time required is limited, even in the case of 70 different treatments.

Our method enabled us to estimate the neighbourhood effect on hospitalized fractures involving the elderly, considering different geographical partitions (10 districts, 23 smaller areas, and 94 more circumscribed zones) without any selection bias due to the different composition of the neighbourhoods. This information will be useful to the Piedmont Region's Epidemiological Service when implementing prevention policies for Turin's population and urban interventions focusing on the neighbourhoods at greatest risk.

The choice of geographical scale is a very important issue in neighbourhood studies, and several authors have suggested considering different scales, and examining neighbourhood effects on outcomes for individuals in more detail, in order to better discern which geographical scale is causally relevant to health (Arcaya *et al.*, 2016). The importance of choosing the most meaningful scale for spatial data is illustrated by a serious analytical issue known as the modifiable areal unit problem (MAUP). Using our MARMoT method, neighbourhood effects can be estimated and compared in different geographical partitions, enabling an assessment of the sensitivity of neighbourhood effect estimates to different choices of geographical scale.

On the other hand, the MARMoT method is strongly influenced by five fundamental aspects:

- the number of variables considered, which directly affects the number of AR values (the number of rows of the table);
- the number of the levels of ordinal and categorical variables and the inclusion of a continuous variable that may increase the entropy of the poset (and the number of rows of the table);

- the number of treatments, i.e. the number of columns in the table; and
- the size of the total population, N, which corresponds to $N = \sum_{r=i}^{R} \sum_{t=i}^{K} f_{r,t}$;
- the choice of f_r that affects both the final dimension of the balanced dataset and the quality of matches.

An increase of one of the first three (rows and columns) without a proportional growth of the fourth point will cause an increase of not exact matching cases with a consequent slight worsening of the balancing. An interesting development would be to test limits of this method changing the first four points in the list. Indeed, when the considered number of treatments is highest, there is still room for improvement. Several adjustments may be done tuning some choices such as the choice of frequency reference f_r for each AR, the caliper and some additional cleaning of less frequent AR values. Further steps will be taken in these direction to improve this already promising technique.

Conclusions

Two main methods were used in the thesis in order to estimate the neighbourhood effect, both are based on propensity score techniques. From both empirical applications and simulations studies it was possible to see that the use of propensity score techniques helps reducing the bias of estimates, thanks to the improvement of the balance of confounders.

However, applications of propensity score techniques in multi-treatment framework with a high number of treatments is particularly challenging, especially, when the number of treatments is high (more than 10), because of computational charge.

Our original proposal, the Matching on Poset based Average Rank for Multiple Treatments (MARMoT) has shown to be promising and useful: it handles high number of treatments, improves the balance and it is pretty fast.

Indeed, the MARMoT proposal solves the curse of dimensionality by summarizing individual characteristics with the help of poset theory. Thanks to poset theory, we are able to assign to each profile (i.e. each combination of individual characteristics) an approximation of its reciprocal position in the ranking of all profiles. Having a unique score that summarizes individuals' characteristics, it is possible to proceed with a matching that assigns to each individual in each neighbourhood one individual in all the other neighbourhoods, and discards those who cannot be matched in order to respect the overlap condition and to make all neighbourhoods comparable. The main idea that underlies this technique is to obtain a population in which each profile is equally represented in all the treatment groups. Having a balanced population, the last step consists in the estimation of the treatment effect using a causal estimand, such as the average treatment effect on treated individuals.

We used this technique to evaluate neighbourhood effect on hospitalized fractures in three different geographical partitions of the city of Turin, 10 districts, 23 areas and 94 zones. The balance of confounders after matching was significantly improved and the algorithm performs the matching very quickly.

Our future work may go in different directions. With respect to the IPTW method presented in chapter 4, some more steps could be taken in using weights derived from other machine learning algorithms. Indeed, the boosting procedure is particularly time consuming, but it is possible to make use of good properties such as flexibility in estimating the functional relation between the treatment and covariates also with other machine learning techniques that are able to estimate the propensity score faster.

Moreover, as far as the MARMoT approach is concerned, it will be necessary to test its sensitivity in different conditions: considering various sets of variables (including eventually continuous, categorical and ordinal variables with many levels), several numbers of treatments and different size of the whole population. Moreover, some choices such as the definition of the frequency reference f_r for each AR, the caliper and some additional cleaning of less frequent AR values should be reconsidered in order to improve the MARMoT procedure. In the MARMoT approach poset based average rank is exploited as a balancing tool with the purpose of reducing the dimensionality. However, it would be interesting to compare its performance with other more common propensity score techniques, such as the propensity score matching, even in a framework with a small number of treatments.

Appendix A

A.1 Parameters to simulate the three scenarios

TABLE A.1: Parameters used to simulate the first scenario.

Variables		Neighbourhoods 1	2	3	4	5	7	8	9	10
(Intercept)	β_0	$-3,1640$	$-0,6430$	$-1,650$	$-2,1240$	$-0,0420$	$-1,6450$	$-2,7530$	$-1,4290$	$0,5070$
Gender	β_1	$0,4420$	$0,1530$	$0,2410$	$0,2640$	$0,0140$	$0,1900$	$0,2980$	$0,1330$	$-0,1350$
Lower Secondary Educ.	β_2	$0,7700$	$0,5770$	$0,5220$	$0,4930$	$0,0270$	$0,3320$	$0,6670$	$0,4550$	$-0,1300$
Upper Secondary Educ.	β_3	$2,1240$	$1,3270$	$1,3030$	$1,3350$	$0,0760$	$0,9540$	$1,8080$	$0,9480$	$-0,2360$
Tertiary Educ.	β_4	$3,0430$	$1,3690$	$1,4980$	$1,5460$	$-0,1280$	$1,3270$	$2,5410$	$1,0460$	$-0,2860$
Hypertension	β_5	$-0,0910$	$-0,0820$	$-0,0820$	$-0,0910$	$-0,0770$	$-0,0200$	$-0,0910$	$-0,0570$	$-0,0390$
Age	β_6	$0,0380$	$0,0100$	$0,0240$	$0,0250$	$0,0030$	$0,0190$	$0,0300$	$0,0140$	$-0,0130$
Overcrowding	β_7	$-0,5490$	$-0,2530$	$0,0640$	$0,0370$	$0,0620$	$0,0020$	$-0,5560$	$0,0260$	$-0,4900$
Drugs	β_8	$-0,0720$	$-0,0190$	$-0,0370$	$-0,0340$	$-0,0020$	$-0,0310$	$-0,0490$	$-0,0160$	$0,0100$

TABLE A.2: Parameters used to simulate the second scenario.

Variables		Neighbourhoods 1	2	3	4	5	7	8	9	10
(Intercept)	β_0	$2,8360$	$-1,0650$	$-0,6010$	$-0,8630$	$-0,9300$	$-0,2560$	$0,8170$	$-0,4910$	$-1,3490$
Gender	β_1	$0,3390$	$1,2150$	$0,3480$	$-0,1640$	$0,1900$	$-0,1910$	$0,5620$	$0,4530$	$2,6710$
Lower Secondary Educ.	β_2	$0,6120$	$0,4030$	$0,3750$	$0,3100$	$-0,1240$	$0,1800$	$0,3300$	$0,1930$	$-0,2680$
Upper Secondary Educ.	β_3	$1,5400$	$0,9330$	$0,9470$	$0,9850$	$-0,2060$	$0,5550$	$1,1300$	$0,3670$	$-0,2740$
Tertiary Educ.	β_4	$2,7400$	$1,3960$	$1,5980$	$1,5920$	$-0,3520$	$1,3990$	$2,0700$	$0,8550$	$0,5070$
Hypertension	β_5	$-0,2960$	$-0,0730$	$0,5710$	$1,4020$	$0,1900$	$0,7710$	$0,2990$	$-0,5720$	$0,5760$
Age	β_6	$-0,1010$	$0,0240$	$0,0060$	$-0,0080$	$0,0150$	$-0,0120$	$-0,0170$	$-0,0170$	$0,0470$
Overcrowding	β_7	$-1,5120$	$-0,1240$	$-0,2110$	$-0,2790$	$0,3490$	$-0,3860$	$-2,1930$	$0,3660$	$-1,7120$
Drugs	β_8	$-0,1870$	$-0,1570$	$-0,1780$	$-0,2260$	$-0,1110$	$-0,1630$	$-0,2530$	$-0,1020$	$-0,1700$
Age^2	β_9	$0,0010$	$0,0000$	$0,0000$	$0,0000$	$0,0000$	$0,0000$	$0,0000$	$0,0000$	$0,0000$
Overcrowding^2	β_{10}	$0,4830$	$-0,1700$	$0,0770$	$0,1970$	$-0,0520$	$0,2150$	$0,6720$	$-0,0840$	$0,5590$
Drugs^2	β_{11}	$0,0190$	$0,0170$	$0,0080$	$0,0180$	$0,0050$	$0,0070$	$0,0140$	$0,0140$	$0,0180$
Gender*Age	β_{12}	$0,0010$	$-0,0090$	$-0,0010$	$0,0110$	$0,0010$	$0,0060$	$-0,0040$	$-0,0030$	$-0,0390$
Gender*Hypertension	β_{13}	$-0,0040$	$0,0800$	$0,0310$	$-0,0310$	$0,0800$	$0,1170$	$0,0530$	$0,0380$	$0,2030$
Gender*Drugs	β_{14}	$0,0010$	$-0,0520$	$-0,0050$	$-0,0130$	$-0,0260$	$-0,0220$	$0,0040$	$0,0010$	$-0,0030$
Age*Hypertension	β_{15}	$0,0080$	$0,0010$	$-0,0090$	$-0,0180$	$-0,0060$	$-0,0090$	$-0,0020$	$0,0080$	$-0,0130$
Age*Drugs	β_{16}	$-0,0010$	$0,0000$	$0,0010$	$0,0000$	$0,0010$	$0,0010$	$0,0010$	$-0,0010$	$-0,0010$
Hypertension * Drugs	β_{17}	$-0,0480$	$-0,0220$	$-0,0160$	$-0,0360$	$-0,0070$	$-0,0360$	$-0,0470$	$-0,0250$	$0,0170$

TABLE A.3: Parameters used to simulate the third scenario.

Variables		Neighbourhoods 1	2	3	4	5	7	8	9	10
(Intercept)	β_0	$-30,6410$	$-0,8200$	$-1,8790$	$-2,5810$	$-0,4830$	$-1,4070$	$-3,3250$	$-1,3680$	$-1,0100$
Gender	β_1	$0,4420$	$0,1530$	$0,2410$	$1,3200$	$0,0140$	$0,1900$	$0,2980$	$0,0266$	$-0,1350$
Lower Secondary Educ.	β_2	$0,7700$	$1,1540$	$0,5220$	$0,2465$	$0,0270$	$0,3320$	$0,6670$	$0,4550$	$-0,1300$
Upper Secondary Educ.	β_3	$2,1240$	$1,3270$	$2,6060$	$1,3350$	$0,0760$	$0,4770$	$1,8080$	$0,9480$	$-0,2360$
Tertiary Educ.	β_4	$3,0430$	$1,3690$	$1,4980$	$1,5460$	$-0,0640$	$1,3270$	$5,0820$	$1,0460$	$-0,2860$
Hypertension	β_5	$-0,0910$	$-0,0820$	$-0,0082$	$-0,0910$	$-0,0770$	$-0,2000$	$-0,0910$	$-0,0570$	$-0,0390$
Age	β_6	$0,3800$	$0,0100$	$0,0240$	$0,0250$	$0,0030$	$0,0190$	$0,0300$	$0,0140$	$-0,0013$
Overcrowding	β_7	$-0,5490$	$-0,2530$	$0,0640$	$0,0370$	$0,6200$	$0,0020$	$-0,0556$	$0,0260$	$-0,4900$
Drugs	β_8	$-0,0072$	$-0,0190$	$-0,0370$	$-0,0340$	$-0,0020$	$-0,0310$	$-0,0490$	$-0,0160$	$0,1000$

A.2 Parameters to simulate the outcome

Table A.4: Parameters used to simulate the outcome.

Variables	**Parameter**	**Value**
(Intercept)	β_0	$-13,553$
Gender	β_1	$0,740$
Lower Secondary Educ.	β_2	$-0,012$
Upper Secondary Educ.	β_3	$0,098$
Tertiary Educ.	β_4	$0,128$
Hypertension	β_5	$0,029$
Age	β_6	$0,100$
Overcrowding	β_7	$0,006$
Drugs	β_8	$0,077$
Neighbourhood 1	β_9	$0,820$
Neighbourhood 2	β_{10}	$1,310$
Neighbourhood 3	β_{11}	$0,375$
Neighbourhood 4	β_{12}	$0,720$
Neighbourhood 5	β_{13}	$0,915$
Neighbourhood 7	β_{14}	$1,430$
Neighbourhood 8	β_{15}	$0,950$
Neighbourhood 9	β_{16}	$1,020$
Neighbourhood 10	β_{17}	$1,535$

Appendix B

B.1 Distibution of individuals in the 10 districts, 23 areas and 94 zones

Table B.1: Distibution of individuals in the 10 districts, 23 areas and 94 zones

Dis.	***23 Areas***	***94 Zones***	***Sect.***	***Sub.***	
1	1 Centro	01-Municipio	75	1409	
		02-Palazzo Reale	9	30	*
		03-Palazzo Carignano	43	830	
		04-Piazza San Carlo - Piazza Carlo Felice	53	825	
		05-Piazza Statuto	68	1481	
		06-Piazza Vittorio Veneto	32	1196	
		07-Borgo Nuovo	26	1021	
		08-Comandi militari	66	1181	
	3 Crocetta	10-Porta Nuova - San Secondo	43	1519	
		18-Politecnico	66	1561	
		26-Crocetta	54	3724	
		27-Ospedale Mauriziano	30	1845	
		28-Corso Lepanto	19	912	
	4 S.Paolo	35-Polo Nord	1	0	
	10 Lingotto	56-Mercati generali	2	32	
	11 S.Rita	54-Stadio Comunale - Piazza d'Armi	8	658	
2	4 S.Paolo	35-Polo Nord	9	1014	
	11 S.Rita	53-Santa Rita	71	5940	
		54-Stadio Comunale - Piazza d'Armi	9	163	
		55-Istituto di Riposo per la Vecchiaia	36	6070	

	12 Mirafiori Nord	59-Corso Siracusa	74	7758	
		60-Fiat Mirafiori	39	3667	
		62-Gerbido	48	5825	
3	4 S.Paolo	34-Monginevro	31	2138	
		35-Polo Nord	54	3215	
	5 Cenisia	17-Porta Susa - Nuovo Tribunale	48	2365	
		17 bis-Carcieri - Officine Ferroviarie	10	219	*
		31-Boringhieri	19	1511	
		33-San Paolo	97	5535	
	13 Pozzo Strada	32-Cenisia	53	4303	
		51-Pozzo Strada	62	5042	
		52-Parco Ruffini - Borgata Lesna	36	4073	
		63-Venchi Unica	82	4671	
4	6 S.Donato	16-San Donato	111	4819	
		25-Teksid - Ospedale Amedeo di Savoia	47	1857	
		29-Campidoglio	80	3882	
	14 Parella	30-La Tesoriera	50	3660	
		50-Parella - Lionetto	85	6712	
		64-Aeronautica	36	1917	
	15 Lucento Vallette	47-Ceronda-Martinetto	9	210	
		49-Parco della Pellerina	5	1	
		65-Le Vallette	2	7	
	16 M.Campagna Lanzo	44-Officine Savigliano	3	0	
5	14 Parella	30-La Tesoriera	4	91	
	15 Lucento Vallette	47-Ceronda-Martinetto	35	1742	
		48-Lucento	88	7199	
		49-Parco della Pellerina	6	378	*
		65-Le Vallette	21	3739	
	16 M.Campagna Lanzo	43-La Fossata	66	4839	
		44-Officine Savigliano	25	1330	
		45-Madonna di Campagna	46	3201	
		46-Barriera di Lanzo	43	1882	
		66-Strada di Lanzo	35	1630	
	17 Borgo Vittoria	42-Borgata Vittoria	46	3762	
		67-Basse di Stura	21	587	*

6	18 Barriera di Milano	36-Cimitero generale	7	225	*
		37-Maddalene	44	3420	
		38-Monterosa	89	5275	
		39-Monte Bianco	41	1684	
	19 Rebaudengo Falchera	41-Barriera di Milano	41	4379	
		68-Barriera di Stura	40	2248	
		76-Villaretto	15	60	*
		77-Falchera	19	2204	
		78-Villaggio Snia - Abbadia di Stura	20	718	
	20 Regio Parco Barca	40-Regio Parco	38	2542	
		79-Bertolla	44	2533	
7	1 Centro	01-Municipio	2	0	
	7 Aurora	12-Borgo Dora	82	2594	
		23-Rossini	51	1445	
		24-Aurora	77	3940	
	8 Vanchiglia	11-Vanchiglia	45	3141	
		21-Gasometro	16	609	*
		22-Vanchiglietta	64	5112	
	18 Barriera di Milano	36-Cimitero generale	14	9	*
	21 Madonna del Pilone	14-Motovelodromo	29	986	
		71-Madonna del Pilone	14	528	*
		72-Sassi	33	750	
		73-Valgrande - Cartman	5	289	*
		74-Val Piana - Val San Martino	7	477	*
		80-Superga	10	128	*
		81-Mongreno	4	52	*
		82-Reaglie - Forni e Goffi	6	231	*
		84-Eremo - Strada di Pecetto	9	143	*
8	2 S. Salvario	09-San Salvario	44	2369	
		09 bis-Parco del Valentino	4	3	*
		19-Piazza Nizza	62	2557	
		20-Corso Dante - Ponte Isabella	80	3859	
	3 Crocetta	28-Corso Lepanto	1	33	
	9 Nizza Millefonti	57-Molinette - Millefonti	3	0	

	22 Cavoretto Borgo Po	13-Parco Michelotti - Borgo Po	64	1543	
		15-Piazza Crimea	34	486	*
		69-Fioccardo	11	587	*
		70-Pilonetto	18	672	
		75-Val Salice	7	138	*
		83-Santa Margherita	7	375	*
		85-San Vito	6	285	*
		86-Parco della Rimembranza	5	42	*
		87-Cavoretto - Val Pattonera	12	456	*
		88-Strada Ronchi - Tetti Gramaglia	8	186	*
9	9 Nizza Millefonti	57-Molinette - Millefonti	70	4263	
		58-Lingotto - Barriera di Nizza	43	3972	
	10 Lingotto	56-Mercati generali	68	5930	
		61-Corso Traiano	70	6564	
10	10 Lingotto	61-Corso Traiano	9	770	
	12 Mirafiori Nord	60-Fiat Mirafiori	2	3	
		62-Gerbido	1	0	
	23 Mirafiori Sud	89-Giardino Colonnetti	42	2539	
		90-Borgata Mirafiori	52	5207	
		91-Drosso	20	2079	
		92-Cimitero Parco Torino sud	3	10	*

B.2 Neighbourhood effect in different geographical partitions (ATT*100), before and after the MARMoT procedure

Table B.2: Neighbourhood effect in different geographical partitions (ATT*100 Before and After the RaMMy procedure)

10	*Districts*		*23*	Areas		94	Zones			
Code	Before	After	Code	Before	After	Code	Before	After	CI	95%
1	0.161	0.044	1	0.12	0.196	01	0.52	0.565	-0.342	1.472
						03	0.182	-0.183	-0.682	0.316
						04	-0.663	-0.506	-1.103	0.091
						05	0.451	0.361	-0.39	1.112
						06	0.017	0.038	-0.557	0.633
						07	0.372	0.106	-0.5	0.713
						08	-0.397	-0.472	-0.828	-0.116
			3	0.138	-0.081	10	0.085	-0.302	-0.67	0.066
						18	0.252	0.293	-0.504	1.09
						26	0.256	0.055	-0.347	0.457
						27	-0.036	0.259	-0.526	1.045
						28	-0.163	-0.2	-0.766	0.366
			11	0.186	0.205	54	0.561	0.361	-0.603	1.325
2	0.038	0.159	4	0.092	-0.043	35	0.164	-0.217	-0.533	0.099
			11	0.186	0.205	53	0.145	0.361	-0.13	0.852
						54	0.561	0.361	-0.603	1.325
						55	0.156	0.123	-0.326	0.572
			12	-0.071	0.062	59	0.026	0.191	-0.228	0.611
						60	-0.059	-0.081	-0.423	0.261
						62	-0.204	-0.047	-0.443	0.35
3	-0.62	-0.061	4	0.092	-0.043	34	-0.062	-0.234	-0.619	0.151
						35	0.164	-0.217	-0.533	0.099
			5	0.025	-0.233	17	-0.143	-0.251	-0.828	0.326
						31	0.09	-0.149	-0.593	0.295
						33	0.075	-0.268	-0.509	-0.027

			13	-0.131	-0.081	32	0.122	0.072	-0.275	0.419
						51	-0.214	-0.081	-0.441	0.279
						52	-0.195	0.225	-0.263	0.713
						63	-0.201	0.174	-0.336	0.684
4	0.026	0.015	6	0.119	0.062	16	0.116	0.31	-0.201	0.822
						25	-0.15	-0.149	-0.606	0.308
						29	0.235	-0.03	-0.397	0.338
			14	-0.072	-0.024	30	0.139	0.157	-0.299	0.614
						50	-0.132	0.004	-0.404	0.412
						64	-0.279	-0.149	-0.661	0.363
			15	0.004	0.186	47	0.226	0.633	-0.155	1.422
5	-0.074	0.11	14	-0.072	-0.024	30	0.139	0.157	-0.299	0.614
			15	0.004	0.186	47	0.226	0.633	-0.155	1.422
						48	-0.23	-0.217	-0.585	0.151
						65	0.331	0.327	-0.25	0.905
			16	-0.139	-0.024	43	-0.015	0.497	-0.12	1.114
						44	0.075	0.004	-0.502	0.51
						45	0.003	-0.2	-0.524	0.125
						46	-0.321	-0.2	-0.711	0.311
						66	-0.724	-0.71	-0.901	-0.519
			17	-0.027	-0.033	42	0.028	-0.013	-0.433	0.407
6	-0.212	-0.104	18	-0.133	-0.205	37	-0.086	-0.132	-0.647	0.383
						38	-0.168	-0.098	-0.539	0.343
						39	-0.072	-0.472	-0.73	-0.214
			19	-0.244	-0.214	41	-0.152	-0.404	-0.64	-0.168
						68	-0.553	-0.251	-0.717	0.215
						77	-0.133	0.191	-0.66	1.042
						78	-0.067	-0.319	-1.002	0.364
			20	-0.235	-0.414	40	-0.277	-0.608	-0.781	-0.435
						79	-0.195	-0.2	-0.582	0.182
7	0.146	0.125	7	0.003	-0.043	12	0.062	0.123	-0.348	0.594
						23	0.415	0.31	-0.307	0.928

						24	-0.196	-0.098	-0.532	0.336
			8	0.297	0.253	11	0.537	0.446	-0.11	1.002
						22	0.117	0.021	-0.362	0.404
			21	0.05	0.015	14	0.519	0.004	-0.517	0.525
						72	0.031	-0.217	-0.765	0.331
8	0.131	-0.152	2	0.165	0.062	09	0.709	0.446	-0.075	0.968
						19	-0.043	0.072	-0.482	0.627
						20	-0.049	0.021	-0.37	0.412
			22	0.066	-0.109	13	-0.061	-0.132	-0.678	0.414
						70	0.736	0.429	-0.541	1.4
9	0.018	-0.028	9	-0.001	-0.005	57	0.084	-0.132	-0.459	0.195
						58	-0.099	-0.047	-0.407	0.313
			10	0.019	0.377	56	0.21	0.31	-0.1	0.721
						61	-0.144	-0.166	-0.541	0.21
10	-0.054	-0.009	10	0.019	0.377	61	-0.144	-0.166	-0.541	0.21
			23	-0.048	0.091	89	0.282	0.514	-0.194	1.223
						90	-0.217	-0.319	-0.575	-0.063
						91	-0.037	0.242	-0.499	0.983

Bibliography

Aaronson, D. (1998) Using sibling data to estimate the impact of neighborhoods on children's educational outcomes. *Journal of Human Resources* pp. 915–946.

Ambrose, A. F., Paul, G. and Hausdorff, J. M. (2013) Risk factors for falls among older adults: a review of the literature. *Maturitas* **75**(1), 51–61.

Antonakis, J., Bendahan, S., Jacquart, P. and Lalive, R. (2010) On making causal claims: A review and recommendations. *The Leadership Quarterly* **21**(6), 1086–1120.

Arcaya, M. C., Tucker-Seeley, R. D., Kim, R., Schnake-Mahl, A., So, M. and Subramanian, S. V. (2016) Research on neighborhood effects on health in the united states: A systematic review of study characteristics. *Social Science & Medicine* **168**, 16–29.

Arpino, B. and Cannas, M. (2016) Propensity score matching with clustered data. an application to the estimation of the impact of caesarean section on the apgar score. *Statistics in medicine* **35**(12), 2074–2091.

Austin, P. C. (2009) Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in medicine* **28**(25), 3083–3107.

Austin, P. C. (2011a) An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research* **46**(3), 399–424.

Austin, P. C. (2011b) Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical statistics* **10**(2), 150–161.

Barnett, D. W., Barnett, A., Nathan, A., Van Cauwenberg, J. and Cerin, E. (2017) Built environmental correlates of older adultstótal physical activity and walking: a systematic review and meta-analysis. *International Journal of Behavioral Nutrition and Physical Activity* **14**(1), 103.

Boccuzzo, G. and Caperna, G. (2017) Evaluation of life satisfaction in Italy: Proposal of a synthetic measure based on poset theory. In *In Complexity in Society: From Indicators Construction to their Synthesis*, pp. 291–321. Springer.

Bolster, A., Burgess, S., Johnston, R., Jones, K., Propper, C. and Sarker, R. (2006) Neighbourhoods, households and income dynamics: a semi-parametric investigation of neighbourhood effects. *Oxford University Press* .

Brüggemann, R. and Carlsen, L. (2011) An improved estimation of averaged ranks of partial orders. Communications in Mathematical and in Computer Chemistry **65**, 383–414.

Brüggemann, R., Lerche, D. and Sorensen, P. B. (2004) First attempts to relate structures of Hasse diagrams with mutual probabilities. *Order Theory in Environmental Sciences* .

Brüggemann, R. and Patil, G. (2011) Ranking and prioritization with multiple indicators introduction to partial order applications.

Buck, N. (2001) Identifying neighbourhood effects on social exclusion. *Urban studies* **38**(12), 2251–2275.

Caceres, I. A., Arcaya, M., Declercq, E., Belanoff, C. M., Janakiraman, V., Cohen, B., Ecker, J., Smith, L. A. and Subramanian, S. (2013) Hospital differences in cesarean deliveries in massachusetts (us) 2004–2006: the case against case-mix artifact. *PloS one* **8**(3), e57817.

Caliendo, M. and Kopeining, S. (2008) Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys* **22**(1), 31–72.

Cannas, M. and Arpino, B. (2018) Machine learning for propensity score matching and weighting: comparing different estimation techniques and assessing different balance diagnostics. RECSM Working Paper Number 54.

Caperna, G. (2016) *Partial Order Theory for Synthetic Indicators*. Ph.D. thesis, University of Padova.

Caperna, G. and Boccuzzo, G. (2018) Use of poset theory with big datasets: A new proposal applied to the analysis of life satisfaction in Italy. *Social Indicators Research* **136**(3), 1071–1088.

Cepeda, M. S., Boston, R., Farrar, J. T. and Strom, B. L. (2003) Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *American journal of epidemiology* **158**(3), 280–287.

Christian, H., Zubrick, S., Foster, S., Giles-Corti, B., Bull, F., Wood, L., Knuiman, M., Brinkman, S., Houghton, S. and Boruff, B. (2015) The influence of the neighborhood physical environment on early child health and development: a review and call for research. *Health & place* **33**, 25–36.

Cochran, W. G. and Rubin, D. B. (1973) Controlling bias in observational studies: A review. *Sankhyā: The Indian Journal of Statistics, Series A* pp. 417–446.

Costa, G., Stroscia, M., Zengarini, N. and Demaria, M. (2017) *40 anni di salute a Torino, spunti per leggere i bisogni e i risultati delle politiche.*

Crane, J. (1991) The epidemic theory of ghettos and neighborhood effects on dropping out and teenage childbearing. *American journal of Sociology* **96**(5), 1226–1259.

Cubbin, C., LeClere, F. B. and Smith, G. S. (2000) Socioeconomic status and injury mortality: individual and neighbourhood determinants. *Journal of Epidemiology & Community Health* **54**(7), 517–524.

Davey, B. A. and Priestley, H. A. (2002) *Introduction to lattices and order.* Cambridge university press.

De Loof, K. (2009) *Efficient computation of rank probabilities in posets.* Ph.D. thesis, Ghent University.

De Loof, K., De Baets, B. and De Meyer, H. (2011) Approximation of average ranks in posets. *Match Commun Math Comput Chem* **66**, 219–229.

Dietz, R. D. (2002) The estimation of neighborhood effects in the social sciences: An interdisciplinary approach. *Social Science Research* **31**(4), 539–575.

Diez Roux, A. V. (2001) Investigating neighborhood and area effects on health. *American journal of public health* **91**(11), 1783–1789.

Diez Roux, A. V. (2004) Estimating the neighborhood health effects: The challenges of casual inference in a complex world. *Social Science & Medicine* **58**, 1953–1960.

Drake, C. (1993) Effects of misspecification of the propensity score on estimators of treatment effect. *Biometrics* **49**(4), 1231–1236.

Dyer, M., A., F. and R., K. (1991) A random polynomial-time algorithm for approximation the volume of convex bodies. *Journal of the ACM* **38**(1), 1–17.

Fattore, M. (2016) Partially ordered sets and the measurement of multidimensional ordinal deprivation. *Social Indicators Research* **128**(2), 835–858.

Galster, G. (2014) Nonlinear and threshold aspects of neighborhood effects. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* **66**(1), 117–133.

Galster, G. and Sharkey, P. (2017) Spatial foundations of inequality: A conceptual model and empirical overview. In *RSF*.

Galster, G. C. (2008) Quantifying the effect of neighbourhood on individuals: Challenges, alternative approaches, and promising directions. *Schmollers jahrbuch* **128**(1), 7–48.

Galster, G. C. (2012) *Neighbourhood effects research: New perspectives.*, chapter The mechanism (s) of neighbourhood effects: Theory, evidence, and policy implications., pp. 23–56. Springer Netherlands, Van Ham, M. et al.

Galster, G. C. and Santiago, A. M. (2006) What's the'hood got to do with it? parental perceptions about how neighborhood mechanisms affect their children. *Journal of Urban Affairs* **28**(3), 201–226.

Garner, L. C. and Raudenbush, S. W. (1991) Neighborhood effects on educational attainment: a multilevel analysis. *Sociology of Education* **64**, 251–262.

Graham, H. (2009) *Understanding health inequalities.* McGraw-Hill Education (UK).

Grannis, R. (1998) The importance of trivial streets: residential streets and residential segregation. *American Journal of Sociology* **103**(6), 1530–1564.

Gu, X. S. and Rosenbaum, P. R. (1993) Comparison of multivariate matching methods: Structures, distances, and algorithms. *Journal of Computational and Graphical Statistics* **2**(4), 405–420.

Hade, E. M. (2012) *Propensity score adjustment in multiple group observational studies: Comparing matching and alternative methods.* Ph.D. thesis, The Ohio State University.

Harding, D. J. (2003) Counterfactual models of neighborhood effects: The effect of neighborhood poverty on dropping out and teenage pregnancy. *American Journal of Sociology* **109**(3), 676–719.

Holland, P. W., Glymour, C. and Granger, C. (1985) Statistics and causal inference. *ETS Research Report Series* **2**.

Hong, G. (2010) Marginal mean weighting through stratification: adjustment for selection bias in multilevel data. journal of educational and behavioral statistics. *Journal of Educational and Behavioral Statistics* **35**(5), 499–531.

Hong, G. (2012) Marginal mean weighting through stratification: A generalized method for evaluating multivalued and multiple treatments with nonexperimental data. *Psychological Methods* **17**(1), 44.

Horvitz, D. G. and Thompson, D. J. (1952) A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association* **47**(260), 663–685.

Iacus, S. M., King, G. and Porro, G. (2012) Causal inference without balance checking: Coarsened exact matching. *Political analysis* **20**(1), 1–24.

Imai, K. and Van Dyk, D. A. (2004) Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association* **99**(476), 854–866.

Imbens, G. W. (2000) The role of the propensity score in estimating dose-response functions. *Biometrika* **87**(3), 706–710.

Imbens, G. W. and Rubin, D. B. (2015) *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Istat (2016) Italia in cifre 2016. http://www.istat.it/it/files/2016/12/ItaliaCifre2016.pdf?title=Italia+in+cifre+2016+-+30%2Fdic%2F2016+-+ItaliaCifre2016.pdf ,
Last access: 27/09/2018.

Kim, D. (2008) Blues from the neighborhood? neighborhood characteristics and depression. *Epidemiologic reviews* **30**, 101–117.

Kwan, M. (2012) The uncertain geographic context problem. *Annals of the Association of American Geographers* **102**(5), 958–968.

Lechner, M. (2001) Identification and estimation of causal effects of multiple treatments under the conditional independence assumption. In *Econometric evaluation of labour market policies*, pp. 43–58. Springer.

Lechner, M. (2002) Program heterogeneity and propensity score matching: An application to the evaluation of active labor market policies. *Review of Economics and Statistics* **84**(2), 205–220.

Leckie, G. (2009) The complexity of school and neighbourhood effects and movements of pupils on school differences in models of educational achievement. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **172**(3), 537–554.

Lee, B. K., Lessler, J. and Stuart, E. A. (2010) Improving propensity score weighting using machine learning. *Statistics in medicine* **29**(3), 337–346.

Lee, B. K., Lessler, J. and Stuart, E. A. (2011) Weight trimming and propensity score weighting. *PloS one* **6**(3).

Lerche, D. and Sorensen, P. (2003) Evaluation of the ranking probabilities for partial orders based on random linear extensions. *Chemosphere* **53**, 981–992.

Li, F., Zaslavsky, A. M. and Landrum, M. B. (2013) Propensity score weighting with multilevel data. *Statistics in medicine* **32**(19), 3373–3387.

Linden, A., Uysal, S. D., Ryan, A. and Adams, J. L. (2016) Estimating causal effects for multivalued treatments: a comparison of approaches. *Statistics in medicine* **35**(4), 534–552.

Linden, A. and Yarnold, P. R. (2016) Combining machine learning and propensity score weighting to estimate causal effects in multivalued treatments. *Journal of evaluation in clinical practice* **22**(6), 875–885.

Lopez, M. J. and Gutman, R. (2017) Estimation of causal effects with multiple treatments: a review and new ideas. *Statistical Science* **32**(3), 432–454.

Lunt, M. (2013) Selecting an appropriate caliper can be essential for achieving good balance with propensity score matching. *American journal of epidemiology* **179**(2), 226–235.

Mair, C. F., Roux, A. V. D. and Galea, S. (2008) Are neighborhood characteristics associated with depressive symptoms? a critical review. *Journal of Epidemiology & Community Health* **62**, 940–946.

Manley, D. (2014) Scale, aggregation, and the modifiable areal unit problem. In *Handbook of regional science*, pp. 1157–1171. Springer.

Manley, D. and van Ham, M. (2012) Neighbourhood effects, housing tenure, and individual employment outcomes. In *Neighbourhood effects research: New perspectives*, pp. 147–173. Springer.

Manski, C. F. (1993) Identification of endogenous social effects: The reflection problem. *The review of economic studies* **60**(3), 531–542.

Manski, C. F. (1995) *Identification problems in the social sciences*. Harvard University Press.

McCaffrey, D. F., Griffin, B. A., Almirall, D., Slaughter, M. E., Ramchand, R. and Burgette, L. F. (2013) A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Statistics in medicine* **32**(19), 3388–3414.

McCaffrey, D. F., Ridgeway, G. and Morral, A. R. (2004) Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological methods* **9**(4).

Meijer, M., Rohl, J., Bloomfield, K. and Grittner, U. (2012) Do neighborhoods affect individual mortality? a systematic review and meta-analysis of multilevel studies. *Social science & medicine* **74**(8), 1204–1212.

Melis, G., Gelormino, E., Marra, G., Ferracin, E. and Costa, G. (2015) The effects of the urban built environment on mental health: A cohort study in a large northern italian city. *International journal of environmental research and public health* **12**(11), 14898–14915.

Merlo, J., Chaix, B., Yang, M., Lynch, J. and Råstam, L. (2005) A brief conceptual tutorial of multilevel analysis in social epidemiology: linking the statistical concept of clustering to the idea of contextual phenomenon. *Journal of Epidemiology & Community Health* **59**(6), 443–449.

Morgan, S. L. and Todd, J. J. (2008) A diagnostic routine for the detection of consequential heterogeneity of causal effects. *Sociological Methodology* **38**(1), 231–281.

Oakes, J. M. (2004) The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. *Social science & medicine* **58**(10), 1929–1952.

Pickett, K. and Pearl, M. (2001) Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. *Journal of Epidemiology & Community Health* **55**(2), 111–122.

Rabe, B. and Taylor, M. (2010) Residential mobility, quality of neighbourhood and life course events. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **173.**(3), 531–555.

Rassen, J. A., Solomon, D. H., Glynn, R. J. and Schneeweiss, S. (2011) Simultaneously assessing intended and unintended treatment effects of multiple treatment options: a pragmatic "matrix design". *Pharmacoepidemiology and drug safety* **20**(7), 675–683.

Ridgeway, G., McCaffrey, D., Morral, A., Burgette, L. and Griffin, B. A. (2006) Toolkit for weighting and analysis of nonequivalent groups: A tutorial for the twang package. *Santa Monica, CA: RAND Corporation* .

Rosenbaum, P. (2002) *Observational Studies.* Second edition. Springer.

Rosenbaum, P. R. and Rubin, D. B. (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**(1), 41–55.

Rosenbaum, P. R. and Rubin, D. B. (1984) Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American statistical Association* **79**(387), 516–524.

Rosenbaum, P. R. and Rubin, D. B. (1985) Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician* **39**(1), 33–38.

Roux, A. V. D., Borrell, L. N., Haan, M., Jackson, S. A. and Schultz, R. (2004) Neighbourhood environments and mortality in an elderly cohort: results from the cardiovascular health study. *Journal of Epidemiology & Community Health* **58**(11), 917–923.

Rubin, D. B. (1976) Inference and missing data. *Biometrika* **63**(3), 581–592.

Sampson, R. J., Raudenbush, S. W. and Earls, F. (1997) Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science* **277**(5328), 918–924.

Sànchez-Riera, L., Wilson, N., Kamalaraj, N., Nolla, J. M., Kok, C., Li, Y., Macara, M., Norman, R., Chen, J. S., Smith, E. *et al.* (2010) Osteoporosis and fragility fractures. *Best practice & research Clinical rheumatology* **24**(6), 793–810.

Sellström, E. and Bremberg, S. (2006) The significance of neighbourhood context to child and adolescent health and well-being: a systematic review of multilevel studies. *Scandinavian journal of public health* **34**(5), 544–554.

Setoguchi, S., Schneeweiss, S., Brookhart, M. A. and Glynn, R. J., . C. E. F. (2008) Evaluating uses of data mining techniques in propensity score estimation: a simulation study. *Pharmacoepidemiology and drug safety* **17**(6), 546–555.

Silber, J. H., Rosenbaum, P. R., Ross, R. N., Ludwig, J. M., Wang, W., Niknam, B. A., Mukherjee, N., Saynisch, P. A., Even-Shoshan, O., Kelz, R. R. *et al.* (2014) Template matching for auditing hospital cost and quality. *Health services research* **49**(5), 1446–1474.

Small, M. L. and Feldman, J. (2012) *Ethnographic evidence, heterogeneity, and neighbourhood effects after moving to opportunity*, pp. 57–77. Springer.

Spreeuwenberg, M. D., Bartak, A., Croon, M. A., Hagenaars, J. A., Busschbach, J. J. V., Andrea, H., Twisk, J. and Stijnen, T. (2010) The multiple propensity score as control for bias in the comparison of more than two treatment arms: an introduction from a case study in mental health. *Medical care* pp. 166–174.

Truong, K. D. and Ma, S. (2006) A systematic review of relations between neighborhoods and mental health. *The journal of mental health policy and economics* **9**(3), 137–154.

Turrell, G., Hewitt, B., Haynes, M., Nathan, A. and Giles-Corti, B. (2014) Change in walking for transport: a longitudinal study of the influence of neighbourhood disadvantage and individual-level socioeconomic position in mid-aged adults. *International Journal of Behavioral Nutrition and Physical Activity* **11**(1), 151.

Uysal, S. D. (2015) Doubly robust estimation of causal effects with multivalued treatments: an application to the returns to schooling. *Journal of Applied Econometrics* **30**(5), 763–786.

de Vuijst, E., van Ham, M. and Kleinhans, R. (2016) A life course approach to understanding neighbourhood effects. *IZA Discussion Paper* **10276**. Available at SSRN: https://ssrn.com/abstract=2861021.

WHO, W. H. O. *et al.* (2011) Global health and aging. *Geneva: World Health Organization* pp. 1–32.

Yen, I. H., Michael, Y. L. and Perdue, L. (2009) Neighborhood environment in studies of health of older adults: a systematic review. *American journal of preventive medicine* **37**(5), 455–463.